Department of Mechanical, Materials and Manufacturing Engineering

UNDERSTANDING SERENDIPITY AND ITS APPLICATION IN THE CONTEXT OF INFORMATION SCIENCE AND TECHNOLOGY

By

XIAOSONG ZHOU

Thesis submitted to the University of Nottingham for the degree of

Doctor of Philosophy

MARCH 2018

To my parents

Abstract

Serendipity is widely experienced in current society, especially in the digital world. According to the Oxford Concise English Dictionary, the term "serendipity" is defined as "the occurrence and development of events by chance in a happy or beneficial way". This PhD research project aims to understand serendipity in the context of information research, and then attempts to design information technologies which can support the encountering of serendipity in cyberspace.

The whole PhD project is organised with two parts. The first part investigates the nature of serendipity by conducting three user studies. After a systematic literature review on existing empirical studies of serendipity, the author finds there are research methodological problems in current studies; for example, the most widely used methods are those conventional ones like interview or survey, and it is mainly the subjective data that can be collected from participants. The author then conducted the first user study, which was an expert interview, where nine experts in the research area of serendipity were interviewed with a focus on the research methodological issues. This study successfully helped the author to gain a broader understanding of the advantages and disadvantages of employing different research methods in studying serendipity. Then the second user study, which was a diary-based study, was performed among a group of Chinese scholars with the aim to have a further investigation on the role of "context" played in the process of serendipity. The study lasted two weeks and successfully collected 62 serendipitous cases from 16 participants. The outcome of this study helped us with a better understanding of how these Chinese scholars experience serendipity, and a context-based research model

was constructed, where the role of external context, social context and internal context were identified in detail during the process of serendipity. One interesting finding from the second user study is that emotions played a role in these participants' experiencing serendipity, which was a part largely ignored by current serendipity researchers; therefore, the author conducted the third user study with the main objective to find out the impact of emotions during serendipitous encountering. This study first employed electrodermal activity (EDA) device to test participants' psychological signals during the process of serendipity, which was implemented through a self-developed algorithm and the algorithm was embedded through a "Wizard of Oz" approach in a sketch game. The results from the study show that participants are more possible to experience serendipity under the influence of positive emotions and/or with skin conductance responses (SCRs).

The second part of the PhD project is the application of serendipity through recommendation technology. Recommender system is an important area that practises serendipity in the digital world, as users in today's society are no longer satisfied with "accurate" recommendations, and they aim to be recommended with the information that are more serendipitous and interesting to them. However, a review of existing studies on serendipitous recommendation, I have found that the inspiring achievements of understanding the nature of serendipity from information science failed to gain attentions by researchers in the area of recommender systems. I then developed a new serendipitous recommendation algorithm by adopting the theory of serendipity from information research, and implemented the algorithm in a real data setting. The algorithm was implemented in Movielens, which involves 138,493 users with about 20,000,263 ratings across 27,278 movies. The evaluation of the algorithm

was conducted in a sub-dataset, which consists of 855,598 ratings from 2,113 users on 10,197 movies. The developed algorithm was compared with another two widely used collaborative filtering algorithms (user-based collaborative filtering and item-based collaborative filtering), and the results demonstrated the developed algorithm are more effective in recommending “unexpected” and “serendipitous” movies to users. A post user study on twelve movie scholars showed that these participants were possible to experience serendipity when they were recommended with movies under the developed algorithm; and compared to user-based collaborative filtering, these participants were more willing to follow the recommended user by the serendipitous algorithm.

Acknowledgement

First and foremost I would like to express my deepest gratitude to my supervisor Dr Xu Sun. She has been supportive since the days I registered as a PhD candidate, and throughout my PhD study, she has continuously supported me by providing constant help and expert guidance. Her constructive criticism, immense knowledge and encouragement made my PhD productive and stimulating. I appreciate all her contributions of time and comments for careful reading and revising the thesis draft in the preparation of this thesis.

In addition, this PhD project is from National Natural Science Foundation of China. I really feel so lucky that my supervisor gives me the chance to conduct this project, which not only gave me financial supports, but also greatly expanded my envision in the research area of human-computer interaction.

I also gratefully acknowledge the full scholarship offered from the University of Nottingham Ningbo China. Their financial support made my PhD work possible.

Last but not the least, I would especially like to thank my parents for their unconditional love, sacrifices, understanding and unwavering support in all my life. In particular, their financial supports and great encouragement sustain me to overcome countless difficulties during my PhD pursuit. Their words can always light me up towards the right direction. Also, thanks go to all the supporters of my PhD study who became a big part of my life. The road is hard working towards a PhD but it is worth striving at the end of the course.

Table of Content

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 Introduction

There is no doubt about the significant role serendipity played in human history. van Andel (1994) collected over a thousand examples of serendipity and highlighted the importance of serendipity from four domains of science, technology, art and daily life (p.637-638):

1) *Science*: X-rays. Röntgen: 'I discovered by chance rays that penetrated black paper.' Years later, when asked, what he then thought: 'I didn't think; I experimented.' (Asimov [1976])
2) *Technology*: Ktesibios, born in Alexandria, a son of a barber, had love and talent for mechanical things. When he hung a mirror in the shop of his father he made a construction by which the mirror could be moved up and down. A counter weight, a ball of lead, hanging invisibly in a pipe, was connected with a rope to the mirror via wheels. When the weight came down the air in the pipe compressed and escaped with a sound. Using this surprising finding Ktesibios was the first to construct hydraulic constructions like a cylinder with a piston.
3) *Art*: One day Picasso had only blue, no other colours. This inspired him to use only blue. The specific effect intrigued him enough to continue with what is now called his 'blue period'. Picasso described his own way of painting as: 'Je ne cherche pas, je trouve.' (I do not search but I found)

4) *Daily life*: Honda introduced large motorcycles in the U.S.A. because analysis of the market had shown the demand for those. The salesmen travelled on small Honda motorcycles. Although the public in the U.S.A. showed a surprising interest in the small ones, it did not occur to the Honda people to sell them: a missed chance. But when the large motorcycles showed defects, the small motorcycles were offered for sale, with great success. Mintzberg [1967] calls this an 'emergent strategy'.

However, based on his substantial analysis of the collected examples, van Andel (1994) argued that "the role of serendipity in science, technology and art is underestimated" (p. 644). Similarly, Buchwald (1999) pointed out that "serendipity appears to have its place in research . . . the process of serendipity should not be underestimated" (p. 4). It was not until the recent three decades that researchers started to recognise the important role of serendipity in scientific research (Erdelze, 2016). For example, Lynch (1999) confirmed the essential value of serendipity in the work of humanities researchers such as historians. Cobbledick (1996) considers serendipity as an important source of artistic stimulation. Senoff (1990) cited personal experience of serendipity occurring within the context of scientific method. Foster and Ford (2003) found serendipity is widely experienced among interdisciplinary researchers.

More recently, the fast developed information technology have made people at every corner of the world to get access with information all around the world. A recent revealed book "Accidental Information Discover — cultivating serendipity in the digital age" (Tammerra and Makri, 2016) highlighted the role of serendipity in current digital age: Serendipity is frequently encountered by users in various digital

environments, such as search and browse environments, social networking environments, next generation library catalogues and web-scale discovery systems.

However, although the study of serendipity has been conducted for three decades (Erdelez, 2016), the work were mainly performed by Western information researchers, and it was not until very recently that Chinese information researchers started to have an eye on this phenomenon. Pan (2010) first introduced the information studies relating to "serendipity" in his master thesis and opened the door of serendipity study to Chinese information researchers. During his thesis, the term "serendipity" was first translated from an academic perspective, and the translation has been accepted by several other Chinese information researchers ever since (e.g. Guo, 2013; Tian & Yu, 2013; Zhou and Huang, 2014). However, these Chinese information researchers still remain on the stage of a theoretical understanding of the existing findings on serendipity from the Western scholars, and to our best knowledge so far, no Chinese scholars have conducted systematic empirical studies to investigate the experience of serendipity among Chinese. This raised our research interest: since current research on serendipity is mainly conducted by Western scholars, whether these research outcomes are also adaptable to Chinese?

On the other hand, with an increasing understanding of serendipity, information researchers started to explore design strategies in digital environments to support serendipity. Makri et al. (2014) reviewed existing design implementations in current digital environments, and have found that there are two main strategies to support serendipity: recommender systems and information visualisation. Information visualisation is concerned with generating interactive, visual representations of information spaces to amplify users' cognition (Card et al., 1999), and it is more of a

technique that can be used as a design strategy to support serendipity; however, in the area of recommender systems, "serendipity" has recently become a hot research topic and garnered much attention (Ge, Delgado-Battenfeld, & Jannach, 2010; Yamaba, Tanoue, Takatsuka, Okazaki, & Tomita, 2013; Sun, Zhang, & Mei, 2013; Adamopoulos & Tuzhilin, 2015; de Gemmis, Lops, Semeraro, & Musto, 2015).

Recommender systems are proved to be a valuable means for coping with the "information overload" problem, which is caused by the explosive growth and variety of information available in the cyberspace (Ricci, Rokach, & Shapira, 2011).The availability of choices on information, instead of producing a benefit, started to decrease users' well-being. Recommender systems are software tools and techniques that can help to provide suggestions for items that to be of use to users (e.g. Resnick & Varian, 1997; Mahmood & Ricci, 2009). Previous research in recommender systems aimed at providing information to users as accurate as possible, and ignored the actual usefulness of suggestions (Herlocker, Konstan, Terveen, & Riedl, 2004). Recent researchers have recognised that factors other than accuracy need to be taken into account to improve the quality of recommendations, and serendipity is one of the factors that have been largely studied and discussed (McNee, Riedl, & Konstan, 2006; Ge, Delgado-Battenfeld, & Jannach, 2010). This raises our another research interest: how technology such as recommender systems can be used in facilitating serendipity?

This thesis aims to investigate serendipitous encountering among Chinese, and the user group of Chinese scholars is selected to conducted empirical studies. This is mainly by a deliberation from two aspects: first, existing studies have demonstrated that serendipity is widely experienced among scholars (e.g. Foster and Ford, 2003; McBirnie, 2008; Sun, Sharples, & Makri, 2011), hence the group of scholars are

possible to provide practical cases of serendipity during the study process, and a recognition and perception on serendipity may help them better experience it in their future research (Makri and Blandford, 2012a); second, the term serendipity was once voted as one of the ten most difficult English words to be translated (Liu, 2013), and its Chinese translation in academic has been accepted and used by Chinese information researchers (Guo, 2013; Tian & Yu, 2013; Zhou and Huang, 2014). Choosing scholars as participants can better accept the academic translation of the term "serendipity", and thus provides convenience for researchers to better explain this phenomenon to participants when carrying out related studies.

1.2 Research Questions

Based on the above discussion, the overall research question throughout the thesis is:

How can information technology help to facilitate Chinese scholar's serendipitous encountering?

This question was further divided into the following four groups of questions:

RQ1: What are existing understandings and implementations of serendipity?

- What is the definition of serendipity in information research?
- How to understand serendipity as part of information behaviour?
- What are the theoretical models of serendipity in information research?
- How serendipity is implemented by current researchers in recommender systems?

The first question group investigates existing theories of serendipity in the context of information research, including definitions of serendipity, understanding serendipity as part of information behaviour, and proposed theoretical research models. An investigation of these theories can help researchers better understand the slippery phenomenon of serendipity, and it is a premise to conduct following studies on Chinese scholars, to further identify whether there are any differences when Chinese scholars experience serendipity.

In addition, an investigation of serendipity study on recommender systems can help information researchers to compare the research of serendipity in areas of information science and recommender systems, and further generate implications to better design systems to support serendipity.

RQ2: What research methods can be employed to understand serendipity?

- What research methods have been employed in current studies of serendipity?
- What are the pros and cons of employing different research methods in studying serendipity?
- What are the challenges when performing serendipitous studies?

While serendipity is considered as a slippery and subjective phenomenon, it is always a challenge to conduct related studies from a methodological perspective (Makri, Bhuiya, Carthy, & Owusu-Bonsu, 2015). A systematic perception on the methodological issues can help me to better design empirical studies of serendipity, pay attention to potential challengeable parts during the study, and also provide

guidance of how to make contributions to existing literatures through a deliberated preparation of research methodology.

RQ3: What is the role of context played in Chinese scholars' experiencing of serendipity?

- How Chinese scholars experience serendipity?
- What are the contextual factors that affect Chinese scholars' serendipitous encountering?
- What is the role of emotion played during the process of serendipitous encountering?

This question group is one of the major investigations of this PhD research. An identification of the role of context played in a serendipitous process can help to understand how Chinese scholars experience serendipity. While existing studies have provided insights of how Western scholars experience serendipity (e.g. Foster and Ford, 2003; McBrine, 2008; Sun et al., 2011), the investigation on Chinese scholars can have an identification whether Chinese scholars also experience serendipity similar to Western scholars. The identified different contextual factors helped me with a deep insight of how the processes of serendipity can be influenced by the surroundings, and in particular, during the study process, I found that emotion played an important role in impacting the process of serendipity, which however is a part that largely ignored by current information researchers. Therefore, I further explored this part of research to extend current understandings of serendipity.

RQ4: How to design recommendation technology to facilitate Chinese scholar's serendipitous encountering?

- How to develop serendipitous algorithm in recommender systems based on the understanding of serendipity in the context of information research?
- How to implement the algorithm in serendipitous recommender system?
- How to evaluate the effectiveness of the serendipitous recommender system?

The fourth question group examined how the findings from information research on the understanding of serendipity can be combined in serendipitous algorithm design in recommender systems, the implementation of the developed algorithm in recommender system through real data setting and the effectiveness of the designed algorithm and recommender system. This part of work provides evidence that through a deliberated design, recommendation technologies can help to facilitate Chinese scholar's serendipitous encountering.

1.3 Research Scope and Outcomes

The thesis was undertaken from the perspective of human computer interaction (HCI), rather than technological development. It embraces user-centred principles by a deep investigation and understanding of how Chinese scholars experience serendipity. The outcomes of this thesis were related to theoretical contributions, methodological contributions as well as recommender system designs. Theoretical

contributions were mainly an extension of the perceptions of serendipity in current literatures; it involves a new context-based theoretical model which depicts the role of context during the process of serendipitous encountering, an investigation of the impact of emotion in experiencing serendipity, and an initial look into the culture dimension in serendipity study. Methodological contributions were concerned with how to conduct cross-cultural empirical studies in understanding serendipity, and how to employ different HCI equipment (e.g. physiological sensors, Eye-tracker) in controlled laboratory settings to observe the occurrence of serendipity. Recommender system designs explored how to combine the perception of serendipity from information research into the design of algorithm, and evaluated the effectiveness of the developed new serendipitous algorithm and system.

1.4 Overview of the Thesis

The structure of this thesis is introduced in Figure 1-1. It begins with a literature review, followed by a review of research methodologies employed in studying serendipity. Next, there are four chapters that describe the main conducted studies throughout the PhD research respectively. The last chapter is the discussion and conclusion of the whole thesis. A more detailed overview of each of the thesis chapters is provided below.

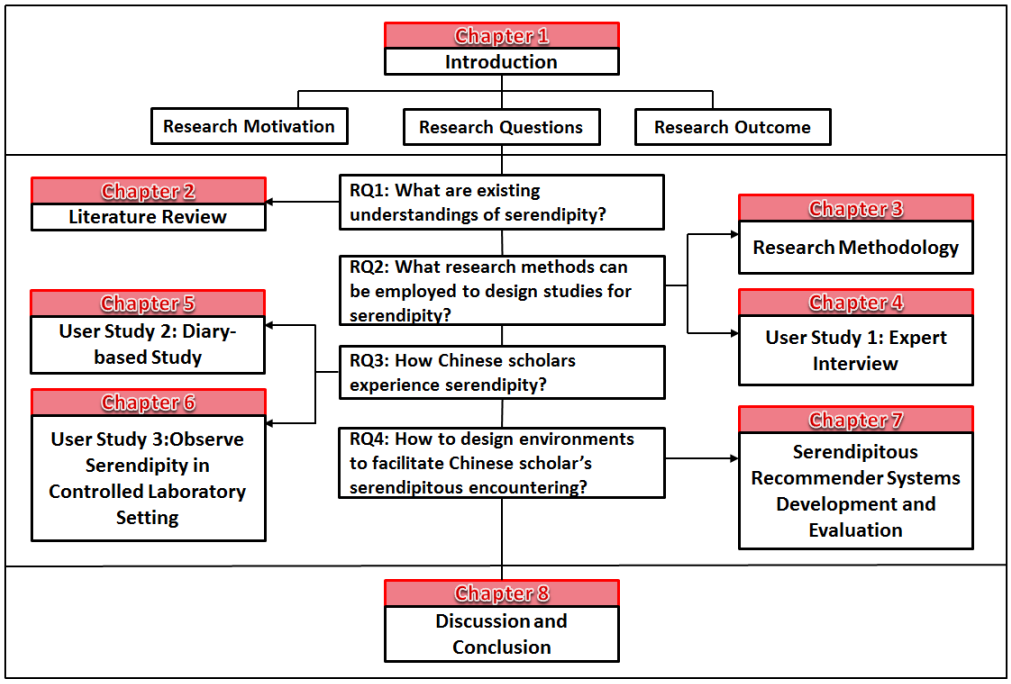


Figure 1-1 Thesis Structure

Chapter 2 — Literature Review

This chapter mainly introduces current study on serendipity from two parts. The first part is Section 2.2, which is an introduction of serendipity in information research, involves understanding serendipity as part of information behaviour in Section 2.2.1, definitions of serendipity given by information researchers in Section 2.2.2, and seven current theoretical models of serendipity in Section 2.2.3. The second part is an introduction of the application of serendipity in current recommender systems in Section 2.3, which introduces approaches of content-based filtering (Section 2.3.1) and collaborative filtering (Section 2.3.2) to provide users with serendipitous recommendations. Section 2.4 is the discussion of this chapter, which lists problems exist in current serendipity studies (Section 2.4.1) and the novel part of this research (Section 2.4.2). Section 2.5 gives the conclusion of this chapter.

Chapter 3 — Research Methodology

This chapter starts by a systematic literature review on existing research methodologies employed in current serendipity studies in the context of information research (Section 3.2), followed with a Section 3.3, which is an brief introduction of all the research methods employed in the following chapters of this thesis.

Chapter 4 — User Study 1 — Expert Interview: An Investigation into Different Research Methods in Studying Serendipity

This chapter introduces a conducted empirical study, which is an expert interview study. Two main objectives were investigated in this chapter: one is the advantages and disadvantages of different research methods in current studies of serendipity, and the other is the confronted methodological challenges in studying serendipity.

Chapter 5 — User Study 2 — Diary-based Study: Understanding How Chinese Scholars Experience Serendipity

This Chapter introduces a diary-based study on Chinese scholars to investigate how they experience serendipity. This study is a follow-up study with previous experience from the similar study conducted in the UK in 2011 (Sun et al., 2011). The study first investigates how Chinese scholars experience study, and then probes into the role of different contextual factors played during the process of serendipity. Finally, with the previous collected data on UK scholars, this study also compared the cultural differences revealed between the two groups of Chinese scholars and the UK scholars.

Chapter 6 — User Study 3 — Observe Serendipity in Controlled Research Setting: the Role of Emotions in Serendipitous Encountering

In Chapter 5, when identifying the role of different contextual factors played during the process of serendipity, an interesting finding revealed is that emotions can also impact the process of serendipity, which however is largely ignored by current information researchers. Therefore, this chapter mainly designed a controlled laboratory study to have a further investigation on the role of emotions in experiencing serendipity. Apart from this main objective, this chapter also paid particular attentions on the methodological parts, as it is not easy to design studies to observe serendipity in controlled laboratory research setting. New research methods are explored to study serendipity, including a new designed platform of a sketch game where a developed serendipitous algorithm is embedded through a Wizard of Oz approach, and the employment of Eye-tracker and physiological sensors in the study.

Chapter 7 — Serendipitous Recommender System Development and Evaluation

This chapter first introduces the new developed information theory-based serendipitous algorithm, which combines the research findings from the understandings of serendipity in Chapter 5. The algorithm is then implemented in real data setting, and two experiments are performed to evaluate the effectiveness of the developed algorithm, one is a vitro experiment performed in the cyberspace and the other is a case study on real users.

Chapter 8 — Discussion and Conclusion

This chapter first introduces the main conclusions of the thesis with respect to the proposed research questions in Section 8.1, and then discusses the contributions of the whole PhD research in Section 8.2. The limitations and future work are introduced in Section 8.3 and Section 8.4 respectively.

CHAPTER 2

LITERATURE REVIEW

Research questions addressed in this chapter:

What are existing understandings and implementations of serendipity?

[1] What is the definition of serendipity in information research?

[2] How to understand serendipity as part of information behaviour?

[3] What are the theoretical models of serendipity in information research?

[4] How serendipity is implemented by current researchers in recommender systems?

2.1 Introduction

This chapter mainly reviews existing literatures on serendipity from two themes: the understanding of serendipity in information research and the study of serendipity in recommender systems. Recent decades have witnessed the increasing studies on serendipity in information research, and a review of the literature in this area can help to gain a better perception on serendipity from a theoretical perspective. In addition, according to Makri et al. (2014), recommender system is one of the major strategies to

support serendipity, and I therefore also had a look into relevant researches in this area. Figure 2-1 gives the overview of this chapter. Section 2.2 reviews studies of serendipity in information research, and it introduces serendipity as part of information behaviour, the definitions of serendipity and the theoretical research models for serendipitous encountering. Section 2.3 reviewed existing serendipitous study in recommender systems, and both content-based recommender systems and collaborative filtering-based recommender systems that are focus on providing serendipitous recommendations are introduced. Section 2.4 gives our discussion and implications based on the literature review, and Section 2.5 is the conclusion of this chapter.

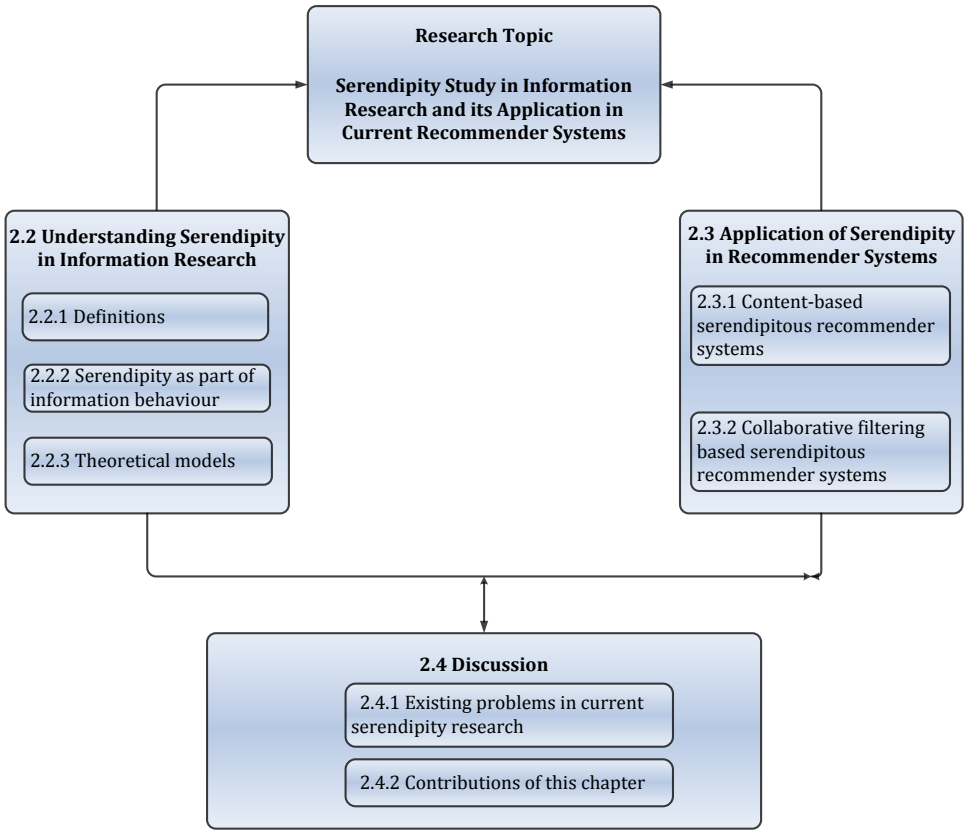


Figure 2-1 Chapter overview

2.2 Understanding Serendipity in Information Research

2.2.1 Definitions of serendipity

The term "serendipity" was coined and created by Horace Walpole on 28 January 1754, in one of his eighteen hundred letters to Horace Mann He knew this British envoy in Florence from his *grand tour* (van Andel, 1994):

> *This discovery indeed is almost of that kind which I call serendipity, a very expressive word, which as I have nothing better to tell you, I shall endeavour o explain to you: you will understand it better by the derivation than by the definition. I once read a silly fairy tale, called The Three Princes of Serendip: as their Highnesses travelled, they were always making discoveries, by accidents& sagacity, of things which they were not in quest of: for instance, one of them discovered that a mule blind in the right eye had travelled the same road lately, because the grass was eaten only on the left side, where it was worse than on the right-now do you understand serendipity?... (you must observe that no discovery of a thing you are looking for, comes under this description)…*

Ever since the term was created, many researchers started to explore the nature of this mysterious phenomenon, and an increasing number of studies are thus carried out. However, till now it seems the definition of "serendipity" has not reached a defaulted one by these researchers. For example, in the Oxford Concise English Dictionary, serendipity is defined as "the occurrence and development of events by chance in a happy or beneficial way". Whereas this definition failed to obtain a consensus among researchers, for instance, van Andel (1994) describes serendipity as "the art of making an unsought finding"; Fine and Deegan (1996) define serendipity as "the unique and contingent mix of insight coupled with chance"; Merton (2004) considers serendipity as "accident and sagacity"; Cooksey (2004) regards serendipity as "the happy convergence of the mind with conditions", while Cunha (2010) take this phenomenon as "the accidental discovery of something that, post hoc, turns out to be valuable".

More recently, McCay-Peet and Toms (2015) conducted a serious of research on serendipity and defined serendipity as "an unexpected experience prompted by an individual's valuable interaction with ideas, information, objects, or phenomena"; based on a review of serendipity in information behaviour, Agarwal (2015) gave his own definition as "an incident-based, unexpected discovery of information leading to an aha! moment when a naturally alert actor is in a passive, non-purposive state or in an active, purposive state, followed by a period of incubation leading to insight and value". Serendipity is also defined by Björneborn (2017) as "what happens when we, in unplanned ways, encounter resources (information, things, people, etc.) that we find interesting".

In spite of these different definitions on the cryptic phenomenon of serendipity, a more general understanding may be implied from the study of Makri and Blandford (2012a). Rather than directly gave a definition on serendipity, they proposed three key elements during a serendipitous encountering, namely "unexpectedness", "insight" and "value". They argued a scenario of serendipity is started by "making a new connection", which involves a mix of unexpectedness and insight, and such connection has the potential to result in a valuable outcome.

2.2.2 Serendipity as information behaviour

There is a long history ever since information researchers have found the special phenomenon of serendipity, and have considered it as part of information behaviour. For example, Apted (1971) pointed out that accidental discovery of information is a special case of browsing that should be distinguished as serendipity. Boyce, Meadow

and Kraft (1994) linked serendipity and browsing: “people find valuable information on subject B when searching for subject A, a phenomenon often called serendipity” (p. 117). Case and Given (2016) argued that “Although most studies of information seeking have chosen to ignore instances of encountering by serendipity, it is obvious that such circumstances are fairly common, especially in online seeking activities (p.106)”.

Erdelez (1997) performed her study on accidental acquisition of information upon 132 information users in an academic environment, and she introduced such kind of information encountering as a type of information behaviour, and it is also won support from other researchers like Williamson (1998) and Toms (2000). She described information users into four categories: super-encounterers, encounterers, occasional encounterers and non-encounterers. Later, Erdelez (2005) gave out the conceptual model of information behaviour related to opportunistic acquisition of information and information encountering, given as the figure below. According to Figure 2-2, opportunistic acquisition of information, also known as serendipity, is on the same hierarchical level as intentional information seeking.

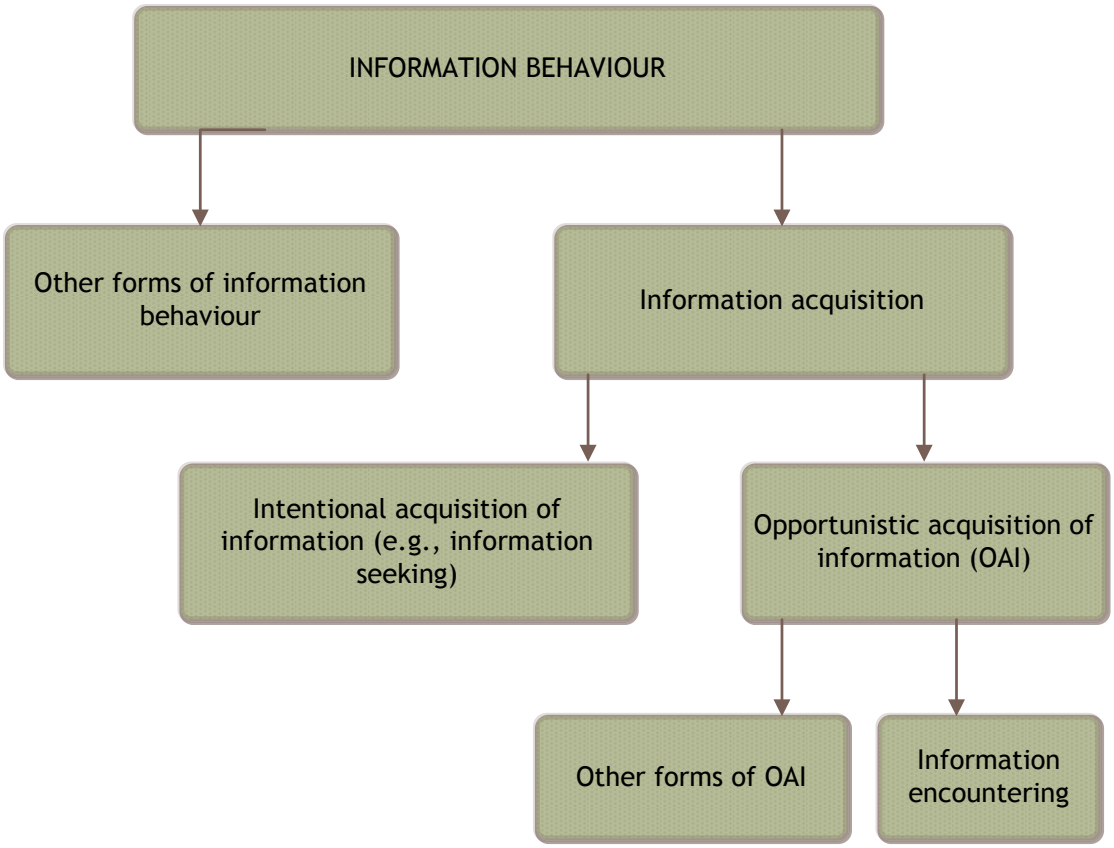


Figure 2-2 Erdelez's (2005) conceptual model of information behaviour

More recently, Agarwal (2015) proposed a nested model (See figure 1), which was extended from Wilson's (1999), where serendipitous information discovery is located within information behaviour, and it should be distinct from the concept of information seeking (usually purposeful), although there might be overlaps between the two concepts. He further discussed that serendipity is possible to occur either in a purposive/active information searching or in a passive/ non-purposive information seeking, and proposed three dimensions for the occurrence of serendipity:

- During a passive phase (including browsing, scanning or non-purposive seeking):

  1. {Accidental, incidental, serendipitous, unintentional or chance}

{encountering, finding, stumbling upon, acquisition or discovery} of information.

- During an active (purposive seeking or search) phase:

  2. {Accidental, incidental, serendipitous, unintentional or chance} {encountering, finding, stumbling upon, acquisition or discovery} of information.

  3. {Opportunistic or active} {encountering, finding, stumbling upon, acquisition or discovery} of information.

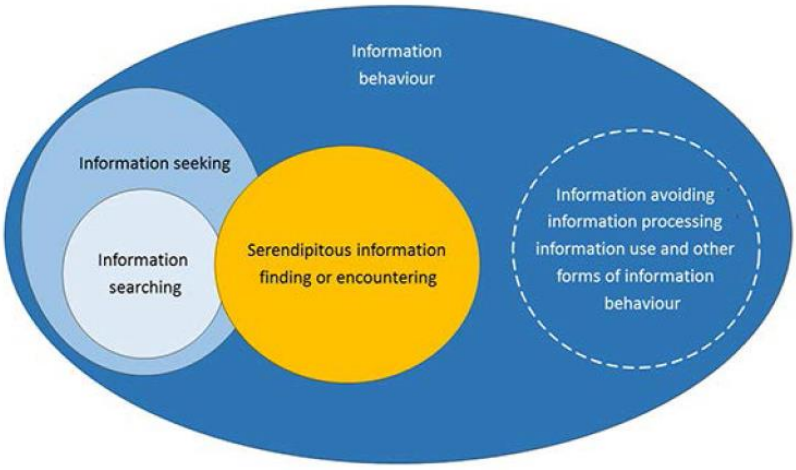


Figure 2-3 Agarwal's (2015) nest model of serendipity

2.2.3 Theoretical models of serendipity

Although the study of serendipity is still an emerging research discipline, several studies in information research have explored how serendipity happens, and theoretical models have been designed by these researchers. Eight existing models for serendipity are reviewed here, five of which are process-oriented models, while the rest two explain the essence of a serendipitous encountering, and the role of context

played during serendipity, respectively.

Process-orientated models

The first model understand serendipity was proposed by Erdelez (2004) who also labelled serendipity as "incidental information encountering" (Figure 2-4). This model focused on the information-seeking process, and the study was undertaken in a controlled environment where users were asked to actively look for information relating to a particular foreground problem but where they actually encountered information relating to a background problem. According to this model, the information encountering process is divided into five stages: *noticing*, *stopping*, *examining*, *capturing* and *returning*. A user's current searching behaviour with regard to the foreground problem is interrupted when he *notices* the information related to the background problem. The user then *stops* to *examine* this information, *captures* any useful details and finally *returns* to the search relating to the foreground problem. This model later won support from Makri and Warwick (2010) in a study of architects' web behaviour.

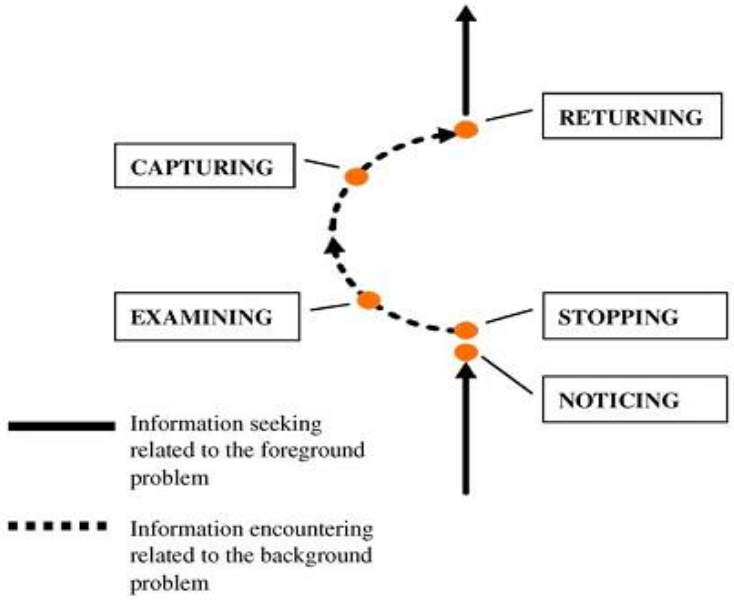


Figure 2- 4: Erdelez's model of serendipity

McCay-Peet and Toms (2010) adapted Cunha's (2005) conceptual model of the serendipity process in organisational management and identified the process of serendipity as follows (Figure 2-5): *while searching for a solution to task A*, with certain *precipitating conditions* a person perceives a *trigger* and then sparks a *bisociation* between disparate, previously unconnected pieces of information and finally this leads to *an unexpected solution to task A or even to a new task B*. The most salient point of this model is the *precipitating condition (Cunha, 2005) which shows* that, to some extent, serendipity can be guided with appropriate strategies.

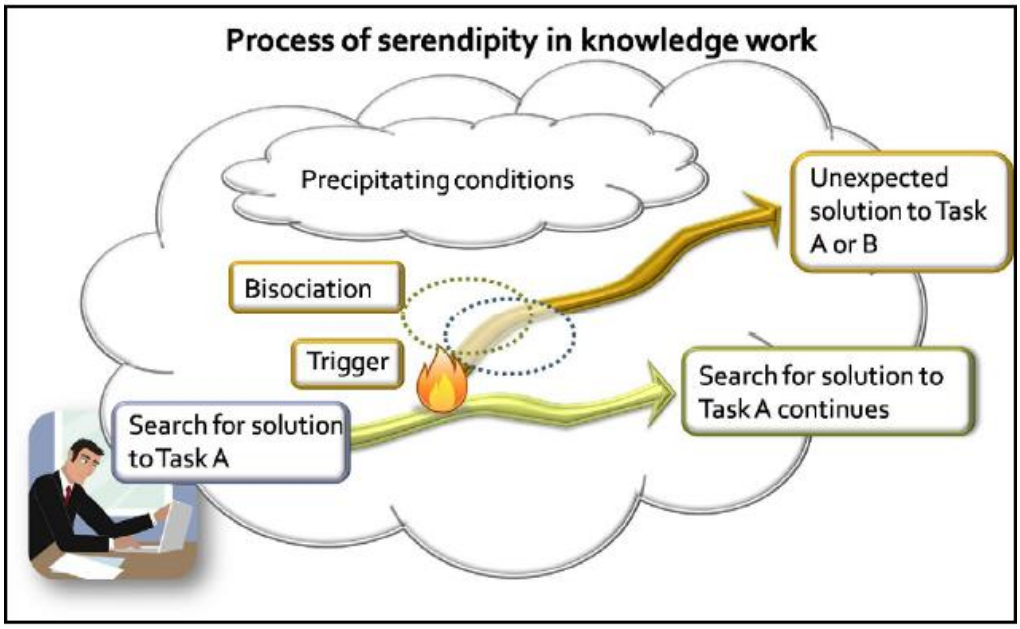


Figure 2-5: McCay-Peet and Toms's model of serendipity

The remaining two models are more focused on the mental processes of individuals who have had serendipitous experiences. Lawley and Tompkins (2008) considered serendipity as "the whole shebang" with six components including the *prepared mind*, an *unexpected event*, *recognised potential*, *seizing the moment*, *amplifying the effects* and *evaluating the effects* (Figure 2-6). They argued that following the removal of any of the six components and the iterative circularity from recognising potential to amplify the effects, the process would no longer be regarded as serendipity. Makri and Blandford (2012) developed their serendipity model based on semi-structured interviews with 28 interdisciplinary researchers. Their findings suggested that *unexpected circumstances* and *insight* could stimulate a person to *make new connections* with an iterative process *by projecting the potential value of an outcome* and further *exploring the value* to gain a *valuable*, *unanticipated outcome (Figure 2-4)*. Based on the discovery, they further proposed a framework, with three

key elements of *unexpectedness*, *insight* and *value*, to classify whether or not an experience can be considered as serendipity.

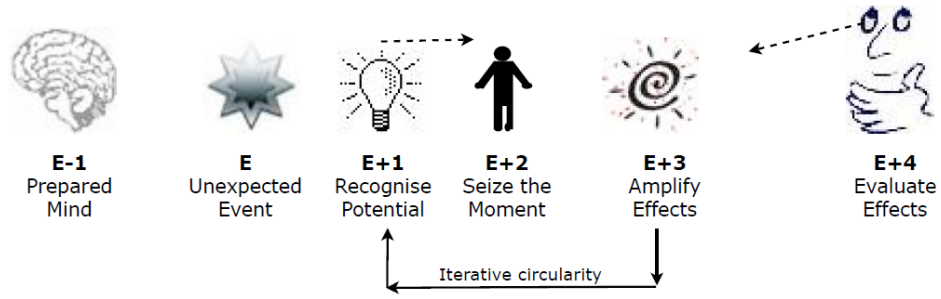


Figure 2-6: Lawley and Tompkins' perceptual model of serendipity

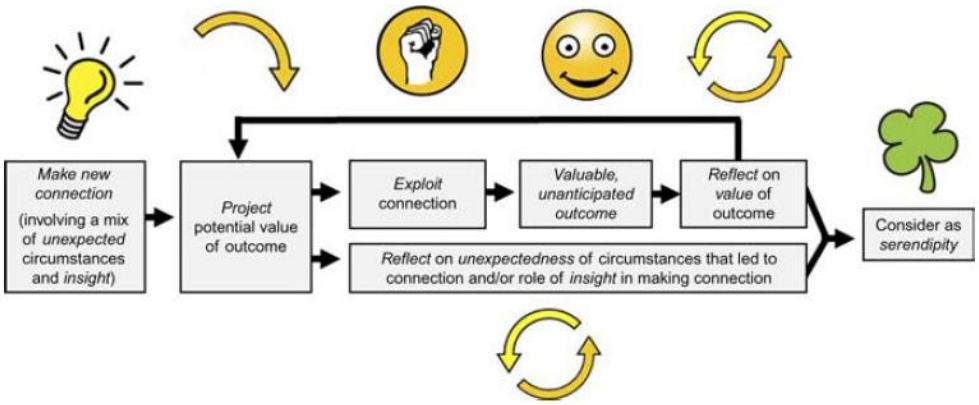


Figure 2-7: Makri and Blandford's perceptual model of serendipity

More recently, McCay-Peet and Toms (2015) updated their serendipity model by combining their previous models with several others (Figure 2-8). They redefined the process of a serendipitous experience as a combination of seven elements, namely, *trigger*, *delay, connection*, *follow-up*, *valuable outcome*, *unexpected thread* and the final *perception of serendipity*. In addition, they argued that unlike other elements, the elements of *delay* and *follow up* "*do not have to happen for perception of serendipity to occur*".

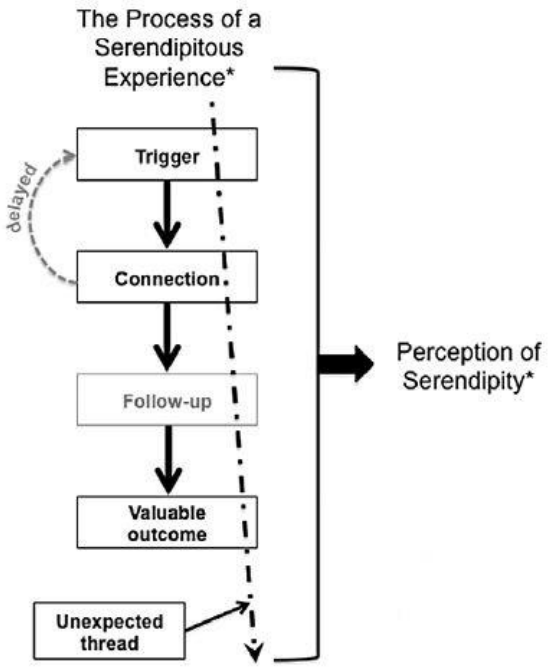


Figure 2-8: McCay-Peet and Tom's updated model of serendipity

Essene-based model of serendipity

Apart from the five process-oriented models, there is another important model which depicts the essence of serendipity (Figure 2-9). Rubin et al. (2011) employed a selective blog minding method by analysing 56 blog entry accounts of chance encounters, from which they identified four key facets which can be used to facilitate serendipity: *a prepared mind* (including a prior concern and previous experience), *an act of noticing* (the ability to notice the provided clue), *chance* (an accidental or unplanned encounter with the find) and *a fortuitous outcome* (unexpected benefits linked to the find). An individual may conclude whether or not an event should be regarded as serendipity by a reframing of these four facets.

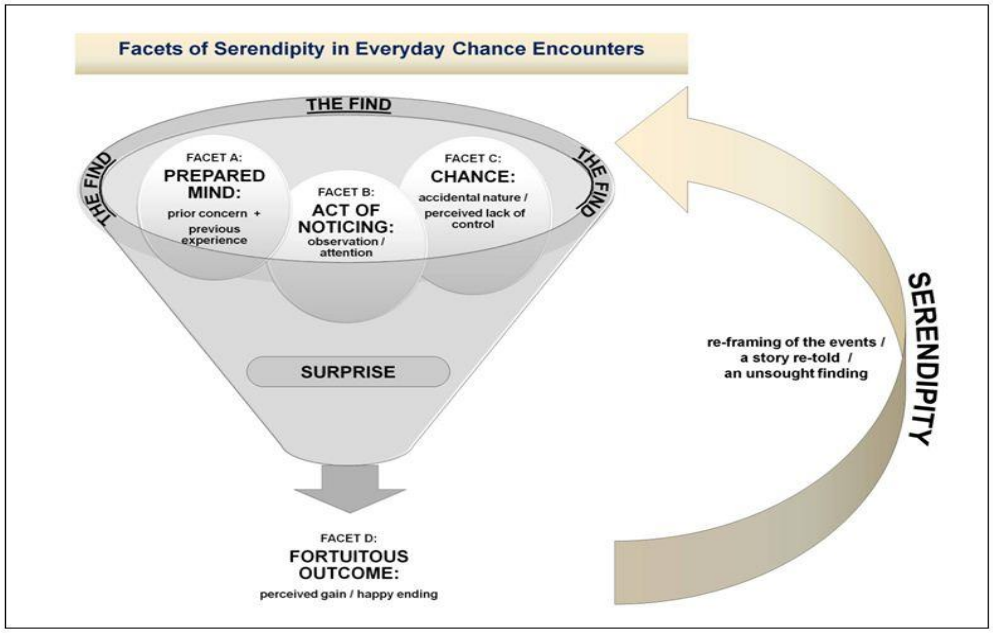


Figure 2-9: Rubin et al.'s conceptual model of serendipity

Context-based model of serendipity

Another model introduces serendipity from the perspective of context. Sun et al. (2011) conducted a mobile-diary study on eleven UK scholars, and have found that context played a role in experiencing serendipity (Figure 2-10). They employed Schmidt's (2000) context model to denote that an individual's readiness to experience serendipity (i.e. *attention*, *pressure* and *focus*) can be influenced by *physical environment*, *social environment* or *temporal factors*. These contextual factors further impact participant's consideration of serendipity from two levels of abstraction, namely, *unexpected finding of information* and *making unexpected connections between information*.

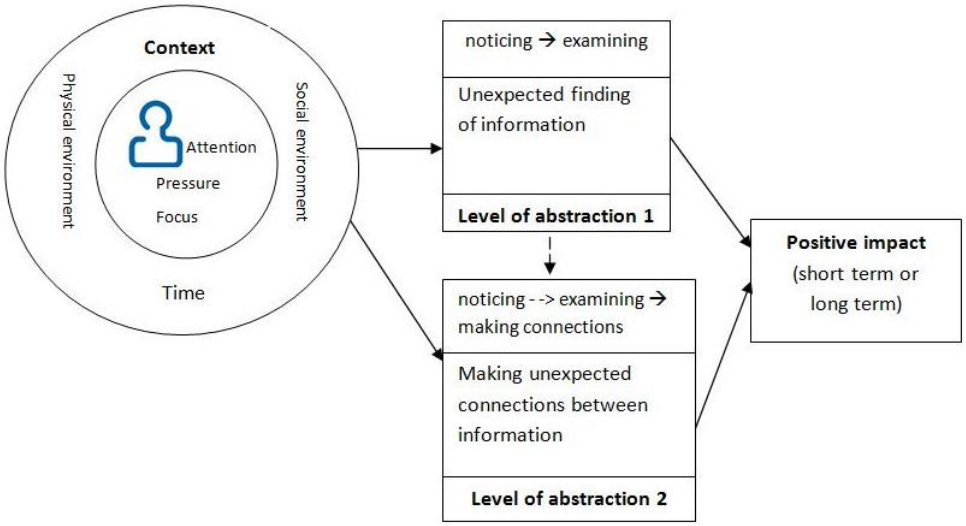


Figure 2-10: Sue et al.'s conceptual model of serendipity

2.3 Application of Serendipity in Recommender Systems

An important area that can implement the design of serendipity is recommender systems. According to Makri et al. (2014), there are two main categories of strategies for the implementation of serendipity: information visualisation and recommender systems. However, serendipity is not an independent research topic in information visualisation; instead this area mainly investigates how data can be transformed to graphical representations by visual layouts or visual interaction methods during information searching, information exploration, knowledge and heterogeneous analysis tasks, to help people with a better cognition and perception of the data (Card et al., 1999; Liu, Cui, Wu, & Liu, 2014; Nazemi, Burkhardt, Hoppe, Nazemi, & Kohlhammer, 2015). Information visualisation is thereby more considered by information researchers as a technical solution of design strategy that can be used in design digital environments that can support serendipity.

However, the situation in recommender systems is quite different. In a recent decade, “serendipity” has also become a popular research topic by researchers in recommender systems (e.g.Yamaba, Tanoue, Takatsuka, Okazaki, & Tomita, 2013; Sun, Zhang, & Mei, 2013; Adamopoulos & Tuzhilin, 2015). “Recommender systems” are considered as filters which suggest items or information that might be interesting to users based on their current information seeking outcomes (de Gemmis, 2015). Conventionally, accuracy, which means the recommended information of the system should be pertinent to a user’s profile, is considered as the most important measurement of recommendation (Herlocker, Konstan, Terveen, & Riedl, 2004); however, with the fast developed IT industry, an increasing number of people start to find that merely recommending accurate information is not enough in current society, and they raise a problem of the so-called “overspecialization” or “serendipity” (McNee, Riedl, & Konstan, 2006). For example, when you searched a movie of “the Lord of the Rings” in the Internet Movie Database (IMDb), it will recommend other movies with similar saga, which are from other science-fiction fans who also liked that movie. Although such recommendation is probably highly relevant to the user, these recommended movies are probably not so useful to users, as they may probably have known these movies. Therefore, researchers in this area start to seek new approaches to improve the user experience in terms of the recommendations, and the idea of “serendipity” have thus gained much attention due to its positive outcome that can bring to users.

Generally, there are two most widely adopted recommendation techniques, namely content-based filtering and collaborative-based filtering. Content-based

recommendation systems mainly implement analysing on a set of user documents, usually textual description, and construct a user profile which is rated by the features of items that an individual user rated previously (Mladenic, 1999). The system then exploits the generated profile and recommends those new items of interest to users. In contrast, collaborative-based recommender system depends on user's opinions. The system collects different user's ratings on objects, and stores these ratings in a distributed database (Resnick, et al., 2009). For instance, the system provides recommendation to user A by computing a neighbourhood user B, whose user profile shares a similar taste to user A. The computing of "similar taste" is based on both user's rating similarity on objects, and the system then recommend user A with those objects that are liked by user B, while have not been rated by A. Researchers in both dimensions of recommender systems have paid attentions on the study of serendipity problems.

2.3.1 Content-based serendipitous recommender system

The pioneering work in this field is performed by Campos & de Figueiredo (2001), who developed MAX, an email-based software agent that simulates user's wandering behaviour on the Internet and recommends users with potential serendipitous URL. A user profile is first generated through the information retrieve techniques on emails, and then a Google™ query is performed on basis of the user profiles with returned URLs. The URLs contains several links that are potentially relevant to user's interests. In the end, a wandering of the agent is performed under the guidance of heuristic functions, and the best-ever visited page address, with

potential serendipitous insights, is sent to the user by email. A pity existing in this study is the empirical results were not so sufficient to present the effectiveness of the agent.

The work of Campos & de Figueiredo provides implications that unexpected information can be recommended to users by determining the "filter bubble" (de Gimmes, 2015) first, and then recommend information that out of the "bubble". Ever since many other researchers start to design the recommendation from this perspective, for example, Kamahara, Asakawa, Shimojo, & Miyahara (2005) employed a similar approach, in which they explored a method to locate unexpected items from clusters similar to a user cluster (the bubble); Abbassi, Amer-Yahia, Lakshmanan, Vassilvitskii, & Yu (2009) developed the Outside-The-Box recommendation, they argued that instead of recommending items that falls into the regions of interest that are familiar to users, serendipitous items can be recommended by those items that are not exposed enough to the user's region.

Iaquinta et al. (2008) used a graph-based methodology to recommend serendipitous information to users. Through a content analyser, they generate a disambiguated document, where a user's interest list is rated according to the semantic analysis of the user documents with the WordNet. They then further specified the text categories into positive ones (user likes) and negative ones (user dislikes). Finally, through a similarity computation of the text categories, those more uncertain items are considered as the unfamiliar ones with most possibilities to be serendipitous and recommended to users.

Oku & Hattori (2011) developed another strategy to support serendipitous

recommendations by mixing different features of two user-items through a fusion-based recommender system, and they further developed the application and tested the system, with a satisfying results returned from the user study (Oku, 2012).

2.3.2 Collaborative filtering based serendipitous recommender system

Collaborative filtering algorithms are also developed to recommend serendipitous information. For example, in the papers published by Kawamae (2010) and Kawamae, Sakano, & Yamada (2009), they proposed a recommendation algorithm that provide serendipitous information to target users by recommending the items listed in the "innovators' preferences". Innovators are considered as those users *"who are aware of items well before their release and purchase these items soon after their release".* Comparing to the other consumers, these innovators are able to search for the latest items in a shorter time, and thus they were ranked with a higher weight by calculating the search time. These higher-weight items, which take into account of the "purchase sequence", also probably match the target user's profile while have not discovered by the target users yet, and is thus considered with potential serendipity.

Lee and Lee (2013) introduced a collaborative-filtering based approach which used dynamically promoted experts among the user pool to produce recommendations. They first calculated item-item similarities, and project the items into an n-dimensional space, so that different clusters of similar items are created, and those users whose profile are concentrated on one cluster, take cluster A as an example, are defined as an "expert" in that cluster, while a user whose profile is not concentrated on cluster A can be considered as a "novice". For a given user, the system first

analysis in which cluster the user is a “novice” and then makes recommendation from “experts” in that cluster to the user. The authors argue that although such recommendations were more focused on novelty, they still remain certain relevance.

Later, Lee and Lee (2015) also applied the concept of entropy and proposed a graph-based recommender system to find recommendations that are both novel and relevant, in other words, those serendipitous items. They identify “positively-related items” that are the set of items in a user’s profile with ratings higher than the user’s average ratings, and these positively-related items play as seed items of recommendations. A graph is then generated by combing the user’s positively-rated items with additional items that were co-rated from other users. The key theory behind their study is the Shannon entropy (Shannon, 2001), which interprets that those items with low entropy are usually considered as accurate and popular recommendations whereas lacking of novelty; in contrast, those high entropy items are less popular but with a higher chance of being novel to recommended users.

In similar vein, Onuma, Tong, & Faloutsos (2009) also applied the graph-based technique and developed the TANGENT recommendation algorithm to recommend surprising, novel but reasonable information to users. The TANGENT takes into account the connectivity to other groups, and selects nodes to connect uses with movies they like. Based on the similarity computing from the graph mining technique, the system gives high scores to those nodes that are well connected to the older choices of the users, while at the same time well connected to unrelated choices. The recommendation strategy expands users’ horizon by recommending items close to user’s current interests, but also undiscovered by the users.

The graph-based technique is also used to employ serendipitous recommendations like mobile apps (Bhandari, Sugiyama, Datta, & Jindal, 2013), music (Taramigkou, Bothos, Christidis, Apostolou, & Mentzas, 2013; Wang et al., 2014), and in User Generated Content contexts (Gobbo, 2007; Yamaba, Tanoue, Takatsuka, Okazaki, & Tomita, 2013; Bordino, Mejova, & Lalmas, 2013). The springing research demonstrates that the study to explore serendipitous application in recommender systems is gaining increasing attentions among researchers in this area.

2.4 Discussion

2.4.1 Existing problems in current serendipity research

A review of the literatures on serendipity study, it can be found that an increasing number of researchers have conducted various studies to uncover the mysterious veil of serendipity in information research and recommender systems. However, it should be noted there are still some problems existing in current studies:

1) *Lack of a culture perspective on understanding serendipity.* According to our review of the existing literatures in information research, it can be found that all of these literatures are discussed based on Westerners as the research participants. However, existing studies have demonstrated that individual's information behaviours can be impacted by culture. For example, Yeh (2007) put forward a model of culture and information behaviour, in which he argued that people from different cultures will develop different traditions, prejudices and habits, and these are the key elements for influencing an individual's information-seeking behaviour. Crow (2011, 2015) conducted studies on information-seeking behaviours for two culture groups, the

American students and the Ugandan students and she found that when compared to the American students, the Ugandan students were more likely to turn to other people for help in their information-seeking quests, whereas the American students were more interested in seeking information independently via creative activities. Therefore, serendipity as part of information behaviour (Erdelez, 1995; Agarwal, 2015), whether or not current research findings – which are mainly generated from Westerners – are also pervasive to other culture groups such as China still needs to be identified.

2) *Existing research models lack a deep look into the role of context played in serendipitous encountering*. Dey et al. (2001) argued that context is necessary "to determine what a user is trying to accomplish. Because the user's objective is difficult to be determined directly, context cues can be used to help infer this information and to inform an application on how to best support users". While information researchers have recognised the important role of serendipity, and proposed different research models to understand serendipity, the context part is largely neglected. The only research studied the role of context was conducted by Sun et al. (2011), where they employed Schmidt's (2000) context model and demonstrated that context can impact an individual's serendipitous encountering. However, their study failed to unpack the context elements in different processes for encountering serendipity, and how these contextual factors influence the processes during a serendipitous experience still remains unaddressed.

3) *A limitation on research methodologies*. Since serendipity is quite a slippery and subjective phenomenon, there are always mythological issues raised for carrying out empirical studies. Makri et al. (2015) argued "serendipity involves an element of

unexpectedness and therefore cannot be created or observed on demand" (p.1), and Foster (2014) argued that "the nature of serendipity raises issues for methodical data collection" (p.18). During my review, it is found that existing studies on serendipity in information research mainly depends on conventional HCI research methods (e.g. interview, survey), which lead to the collected data are mainly subjective based, and lack of objective data. In addition, the collected subjective data in most of studies are retrospective, and depends on participant's memories. There is a lack of research to explore different methods to have a deep look into the instant process of serendipitous encounter. This part will be introduced in detail in next chapter.

4) *Lack combination in the area of information research and recommender systems*. During the review process, I found that although researchers in recommender systems have recognised the importance of serendipitous recommendations for today's users, they are still struggle to identify the different meanings of "novelty", "diversity" and "serendipity" (Ge, Delgado-Battenfeld, & Jannach, 2010; Kawamae, 2010; Taramigkou, Bothos, Christidis, Apostolou, & Mentzas, 2013). An important reason is that the outcomes from information research on the study of serendipity in recent years are not acknowledged by these researchers in recommender systems. For example, Herlocker et al. (2004) describes serendipitous recommendations as the ones helping the user to find surprisingly interesting items she might not have discovered by herself. McNee et al. (2006) identify serendipity as the experience of receiving an unexpected and fortuitous item recommendation, while Shani and Gunawardana (2011) state that serendipity involves a positive emotional response of the user about novel items and measures how surprising these recommendations are. These

researchers then start design different methods and algorithms to make serendipitous recommendations based on their perception of serendipity, either through content-based design strategies or collaborative filtering strategies. However, as Makri et al. (2014) argued, most of these designs and implementations "are based on an intuitive (rather than empirically grounded) understanding of the phenomenon and this makes it difficult for designers of digital information environments to know what functionality it is particularly useful to support" (p.2183). Current design for serendipity in recommender systems heavily depends on designer's perception of serendipity and is mainly from the designer's perspective, and lack empirical perception of how users experience serendipity, which is actually one of the main research points in information research.

2.4.2 Novel part of this research

Based on the discussed problems in current serendipity research, the novel parts of this research are outlined below:

- *Culture*. While current research outcomes are mainly obtained from Westerners, this research takes Chinese scholars as the participant group, and conducted empirical studies to examine how these Chinese scholars experience serendipity. The selected user group makes the research as an initial probe into the role of culture in serendipitous encountering.
- *Context*. This research extended the understanding of the role of context in experience serendipity. Compared with previous context model by Sun et al. (2011), this research further unpacked the contextual factors during the

process of serendipitous encounter, identified different contextual factors that impact an individual's experience of serendipity and constructed an updated context-based research model. The identified contextual factors including external context (e.g. time, location, status), social context (e.g. different social partners) and internal context (i.e. precipitating conditions, sagacity/perceptiveness and emotions).

- *Research methodology.* Through a systematic literature review on exist research methods conducted in understanding serendipity (next chapter introduces the details of this part), this research also explored new approaches to perform serendipity studies. In particular, the HCI equipment of Eye-trackers and physiological sensors were employed to capture the objective data during the process of serendipity. The exploration of these new approaches made it possible for researchers to capture the instant and objective data when serendipity occurs, thus contribute to current research methods which mainly depend on collecting participant's retrospective and subjective data.
- *Bridge the gap between information research and recommender systems on serendipity study.* Another novelty of this research is that it combines the outcomes from information research on understanding serendipity to recommender systems. An information theory-based serendipitous algorithm was developed, which gave new insight of serendipity cross-disciplinarily. The follow-up empirical studies successfully implemented the algorithm in real-datasets, and provide evidence for the effectiveness of the developed

algorithm in making serendipitous recommendations to users.

2.5 Conclusion

This chapter reviewed existing research on serendipity in both areas of information research and recommender systems. As part of information behaviour, the definitions of serendipity and theoretical models of understanding serendipity in information research are introduced, in addition with the implementation of serendipity in current recommender systems. Based on the review, four problems existing in current serendipity research were proposed: lack of a culture perspective on understanding serendipity; existing research models lack a deep look into the role of context played in serendipitous encountering; a limitation on research methodologies; and lack combination in the area of information research and recommender systems. Following these problems, five novelty parts throughout the whole research are discussed: culture, context, research methodology, emotions and bridge the gap between information research and recommender systems on serendipity study.

CHAPTER 3

RESEARCH METHODLOGY

Research questions addressed in this chapter:

What research methods can be employed to design studies for serendipity?

[1] What research methods have been employed in current studies of serendipity?

3.1 Introduction

The previous chapter reviewed theoretical parts of serendipity study in the context of information research, and also discussed its application in recommender systems. However, it should be noted that serendipity *per se* is quite a subjective phenomenon - "it can mean different things to different people, in different situations" (Makri and Blandford, 2012b, p.707), and hence there is always research methodological problems when carrying out serendipity studies:

- *"There is no doubt about the importance and impact of serendipity. However, it is also clear that the nature of serendipity raises issues for methodical data collection" (Foster and Ellis, 2014, p.18)*
- *"Serendipity involves an element of unexpectedness and therefore cannot be created or observed on demand." (Marki et al., 2016, p.1)*

The main objective of this chapter is to review and analysis existing methodologies applied in understanding serendipity in information research. A systematic literature review is conducted in Section 3.2, which investigates methodological issues in current studies of serendipity in the context of information research, and Section 3.3 introduces different research methods applied in empirical studies throughout the thesis. Section 3.4 is the conclusion of this chapter.

3.2 Systematic Literature Review

In order to have a thorough review on those research methods applied on serendipity studies, a systematic literature review is conducted on current studies of serendipity in the context of information research, which focuses to understand the esoteric phenomenon of serendipity. The systematic review excels through a comprehensive search and has focused on the research methodology. The identified studies included in this review were primarily from databases of Web of Science, Scopus and Elsevier. The terms used to perform the search included: serendipity, information encountering, chance encounters, incidental information acquisition, and opportunistic acquisition of information. Overall, 596 potential articles were derived from the search results. All of the articles were evaluated on the basis of the criteria listed in table 1. For the ultimate inclusion in the review, thematic relevance was the most decisive criterion.

Table 3-1 Evaluation Criteria

Publication Language	English
Journal	Only peer-reviewed journals
Author	Articles of popular authors irrespective of ranking, but must be peer-reviewed
Setting	Empirical Studies (e.g. interview, controlled laboratory study)
Research Area	Information Science
Date of publication	2000-2017

During this process, 569 articles were excluded after reading the titles, keywords and abstracts; this is mainly because the following two reasons: 1) these articles are not included in the context of information research; 2) these articles are not discussed in the scope of understanding serendipity, such as how participants experience serendipity, or how serendipity can be facilitated.

27 potential relevant articles are thus remained. After a full text reading of these 27 papers, four articles were excluded (i.e. Agarwal, 2015; Björneborn, 2017; Erdelez, Basic, and Levitov, 2011; Foster and Ellis, 2014) from our discussion because these four papers mainly discuss serendipity on basis of a literature review, and no empirical studies are involved. As a result, 23 journal articles were finally screened to a further discussion in the following sections.

3.2.1 Year of publishing

Figure 3-1 lists the published year for the selected 24 articles which involves empirical studies relating to the research topic of serendipity. It can be found that before the year of 2011, empirical studies relating to serendipity research is fairly less, and the turning point appears in 2011, ever since an increasing number of empirical

studies relating to serendipity are published. A possible reason for such phenomenon is owing to the special issue of "Information Research — on the opportunistic discovery of information" in 2011, and through this special issue, the role of serendipity is highlighted in the context of information research.

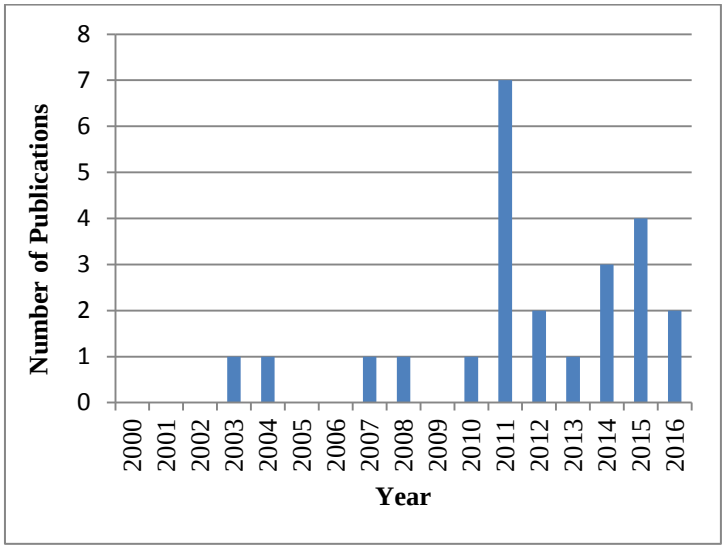


Figure 3-1 Number of published serendipity papers with empirical studies

3.2.2 Employed methodologies

Table 2 lists the methodologies employed in our screened 24 journal articles. It can be found from the table that various HCI research methods have been employed to study serendipity, including interview, survey/questionnaire, observation, think-aloud, diary-based study, controlled laboratory study, online ethnography, Wizard of Oz, Selective blog mining, Narrative and network analysis. There are some empirical studies conducted merely by one method (e.g. interview, questionnaire, selective blog mining), and other studies employed a mixed method to understand serendipity.

Table 3-2 Employed research methods in the identified 24 journal articles

Employed Methodology	Articles	Research Setting	Number
Interview	(Foster and Ford, 2003) (Nutefall and Ryder, 2010) (Pálsdóttir, 2011) (Makri and Blandford, 2012a, 2012b) (Makri, Blandford, Woods, Sharples, and Maxwell, 2014) (Jiang, Liu, and Chi, 2015) (McCay-Peet and Toms, 2015)	Naturalistic	8
Survey/Questionnaire	(Heinström, 2007) (Stewart and Basic, 2014) (McCay-Peet, Toms, and Kelloway, 2015)	Naturalistic	3
Selective blog mining	(Rubin, Burkell, and Quan-Haase, 2011)	Naturalistic	1
Interview and Survey/Questionnaire	(Yadamsuren and Heinström, 2011) (McCay-Peet, Toms, and Kelloway, 2014)	Naturalistic	2
Observation, think-aloud and interview	(Björneborn, 2008)	Naturalistic	1
Survey/Questionnaire	(Erdelez, 2004) (McCay-Peet and Toms, 2011)	Controlled laboratory	2
Think-aloud and interview	(Makri, Bhuiya, Carthy, and Owusu-Bonsu, 2015)	Controlled laboratory	1
Think-aloud, Survey/questionnaire and interview	(Miwa et al., 2011)	Controlled laboratory	1
Diary-based study and interview	(Sun, Sharples, and Makri, 2011) (Kefalidou and Sharples, 2016)	Naturalistic	2
Narrative and network analysis	(McBirnie and Urquhart, 2011)	Naturalistic	1
Online ethnography, Survey/questionnaire and interview	(Saadatmand and Kumpulainen, 2013)	Naturalistic	1
Wizard of Oz, Survey/questionnaire and interview	(Pontis et al., 2016)	Naturalistic	1

Collected data type: qualitative vs. quantitative

As can be seen from the table, most of the data from these empirical studies is qualitative in nature, and among the various research method, interview is the most widely employed one. Eight out of 24 empirical studies (1/3) are conducted merely

based on interview. For example, Foster and Ford (2003) had carried out open-ended interview on academic researchers, and reinterpreted the notion of serendipity in information-seeking contexts. They found serendipity was widely experienced among inter-disciplinary researchers and it played an important role in researcher's information encountering and generating new ideas. Makri and Blandford (2012a) performed semi-structured interviews on interdisciplinary researchers, and based on their memorable serendipitous encountering in their research and daily life, they had put forwarded a memorable process model of serendipity and had further presented a framework to identify serendipity, where "unexpectedness", "insight" and "value" were considered as three factors to evaluate the "strength" of serendipity (Makri and Blandford, 2012b). By the collected cases of online information encountering from the interview, Jiang, Liu and Chi (2015) had drawn an integrated model of online information encountering which is on basis of McBirnie's (2008) consideration of the process-perception duality of serendipity. The model divides information encountering process into pre-, mid- and post-activities, whereas the perception part is constitute with three clusters of user, information and environment, each cluster is made up of constant factors and dynamic factors.

Apart from employing interviews, some empirical studies also used interview together with other methods (e.g. Björneborn, 2008; Yadamsuren and Heinström, 2011;McCay-Peet, Toms, and Kelloway, 2014; Miwa et al., 2011), to collect qualitative data from participants. In addition, there are other research methods that can collect qualitative data more than interview. For example, Björneborn (2008) collected empirical data through a naturalistic observation of users' information behaviours in two physical public libraries, and they also adopted think-aloud

comments and reflections on what triggered user's attention and information behaviour. Saadatmand and Kumpulainen (2013) used online ethnography to investigate participant's serendipitous learning in open online networks in the context of MOOCs, these participants are from different countries around the world and their shared online data (e.g. blogs, Twitter, Facebook and YouTube) are collected to make analysis. In a similar vein, Rubin et al. (2011) used the method what they called as "Selective blog mining" to examine examples of serendipity in everyday life by retrieving and analysing the content from various users shared on GoogleBlog. Sun et al. (2011) and Kefalidou and Sharples (2016) conducted diary studies, where users' considered serendipity cases are collected through mobile applications as the entry of diary. McBirnie and Urquhart (2011) conducted narrative and network analysis to detect motifs of interaction patterns between people and inforamtion in serendipty, which was mainly based on examining fifty narrative data collected from the Citation Classics database.

Comparing to qualitative data, quantative data is mainly collected through survey/questionnaire. In our identified 24 empirical studies, three of them are purely based on survey/questionnaire. Heinström (2006) discussed psychological aspects in serendipity (which they call as incidental information acquisition) through three survey studies covering the age of participants ranging from 12 to 53, and they found that an energetic personality, high motivation, and positive emotionality can enhance the possibility of experiencing serendipity, while low motivation, stress, and insecurity reduced serendipity. Stewart and Basic (2014) surveyed 94 students with their online information encountering experiences and found online information encountering is experienced by these students but the encountered information were

not captured by them for future use using those web-based tools. McCay-Peet and Toms (2015) conducted a web-based survey on 289 participants and found that certain types of digital environments (e.g., websites, databases, search engines, intranets, social media sites) are more conducive to serendipity, and those environments that are trigger-rich, enable connections, and can lead to unexpectedness are more likely to foster serendipity. In addition, seven out of the identified 24 empirical studies employed survey/questionnaire together with other methods to understand information encountering or measure serendipity (i.e. Yadamsuren and Heinström, 2011; McCay-Peet et al., 2014; Erdelez, 2004; McCay-Peet and Toms, 2011; Miwa et al., 2011; Saadatmand and Kumpulainen, 2013; Pontis et al., 2016).

There is another study collected different types of quantitative data. Miwa et al. (2011) conducted a controlled laboratory study during which Eye-tracker is employed to investigate information encountering, and the collected data from eye-tracker involves quantitative data such as number of viewed web pages, number of eye-gaze points, duration time on different web pages, etc. With the collected objective quantitative data, they argue that participants can better recall their feelings and thoughts at the very moment of information encountering, and participants acted differently in well-defined tasks and exploratory searches, which may lead to a "reconsideration of the definition of information encountering".

Naturalistic setting vs. controlled laboratory setting

André et al. (2009) argued that "because serendipity is inherently rare, it is hard for researchers to capture or induce it for study and experimentation" (p.307). Similarly, Björneborn (2017) argued "We cannot design environments always leading

to serendipity – as serendipity is a highly subjective and situational phenomenon" (p.1068). Nevertheless, Erdelez (2004) contended that challenges study serendipity in controlled environments "can be overcome with very careful planning, high attention to detail, and ongoing adjustments in a development and execution of a research design" (p. 1023). While most of current studies on serendipity are conducted in naturalistic research setting, in our identified the 24 empirical studies, besides Miwa et al.'s study, there are still three other relevant studies that are carried out under controlled laboratory setting.

Erdelez (2004) designed a laboratory setting where ten participants were assigned with a web-searching task related to the online shopping for a surfboard as a foreground task, and among the searched results an information encountering trigger was embedded, which is related to their background task (information on Web analytics). After the task is finished, they also conducted a post survey to investigate participants' perceptions on information encountering. The results from the study showed that although their participants failed to fully accomplish an information encountering episode (noticing, stopping, examining, capturing, and returning), they expanded the understanding of "noticing" stage, and proposed possible solutions such as to use eye-trackers to better capture subtle changes during this stage. This study also gave an initial insight into observing participants' reactions to information encountering in a controlled laboratory setting, and demonstrated methodological possibilities to observe serendipity in controlled laboratory research settings.

McCay-Peet and Toms (2011) invited participants to a scenario of office environment, where 123 valid participants were asked to conduct web browse in a wikiSearch system for a period of twenty minutes. The system involves different

interfaces that not only can give feedback to participants of their search tasks, but also provides a Suggested Items list which are somewhat related to the searched result. The participants were then invited to a survey regarding serendipity scales, which was organically drawn from Björneborn's (2008) ten dimensions of the physical library that may support for serendipity. Their study finally identified five factors (enabled connections, introduced the unexpected, presented variety, triggered divergence, and induced curiosity) as the core elements that can facilitate serendipity in digital environments.

Another controlled laboratory study was conducted by Makri et al. (2015). They recruited 45 participants performing self-selected searching tasks in three different digital environments: digital libraries, e-commerce sites and online new sites. These participants were asked to bookmark and/or screenshot the information that they considered as useful in at least 30 minutes, and a think-aloud method was performed during the whole task process. After the tasks were finished, a post interview was conducted to every participant, which focused on understanding two dimensions of "unexpected" and "useful" towards users' stored bookmarks or screenshots. The participants were not received with the "real" research purpose, which was to investigate serendipity, until the end of the study. The successfully employment of the study demonstrates that "with a carefully-considered approach, serendipity-related information interaction behaviour can be directly observed".

3.2.3 Discussion

After an identification of the 24 empirical studies, it can be found that there are two major concerns in current serendipity studies in the context of information

research. The first is that so far it is still the conventional research methods (e.g. interview, survey) that play the main role in understanding serendipity. A disadvantage of employing these research methods is that they are heavily dependent on participants' memories, and the collected data are subjective and retrospective. While such collected data can help to picture participants' general understanding through an already occurred serendipitous encounter, it failed to help researchers to find potential characters or elements during the process of a serendipitous episode. Foster (2014) argues "there must be a way for the researcher to learn when a serendipitous recognition occurs, either at the moment of the event or at some later point" (p.19). New attentions should be paid on this aspect which can help researchers to capture the instant processes when serendipity occurs.

The second concern is about the collected data type. Most of the discussed researches only collected subjective data, and it is therefore heavily rely on individual's subjective perceptions on serendipity. However, as Erdelez (1997) identified from her study, there are individual differences in encountering serendipity, ranging from super-encounterers to non-encoutneres, thereby the perception for different individuals can also be quite different. Other information researchers like Rubin et al. (2011), Makri and Blandford (2012a), and McCay-Peet and Toms (2015) all argued that the recognition of serendipity requires a re-framing of the whole encountering, and it is even a kind of ability for individuals to have the recognition. In other words, it is possible for an individual that even serendipity is happening to him/her, s/he cannot recognise it timely. A possible solution to address this concern is to collect more objective data during the process of serendipity, such as Miwa et al.'s (2011) work that employed Eye-tracker to collect objective data like participants'

number of eye-gaze points, duration time on different web pages, etc. An accumulation of the collected objective data can provide more evidences for researchers to evaluate whether an episode is serendipity or not, to better understand the nature of serendipity and no more rely on participant's subjective memories. In addition, the collected objective data can also help participants with clearer memories on their experiences and thus better recognise serendipity.

3.3 Methods Applied in This Research

Figure 3-2 depicts the research methods applied in this thesis. A literature research is first conducted, followed by three different user studies in the context of information research, and the design and evaluation of the serendipitous algorithm in recommender system.

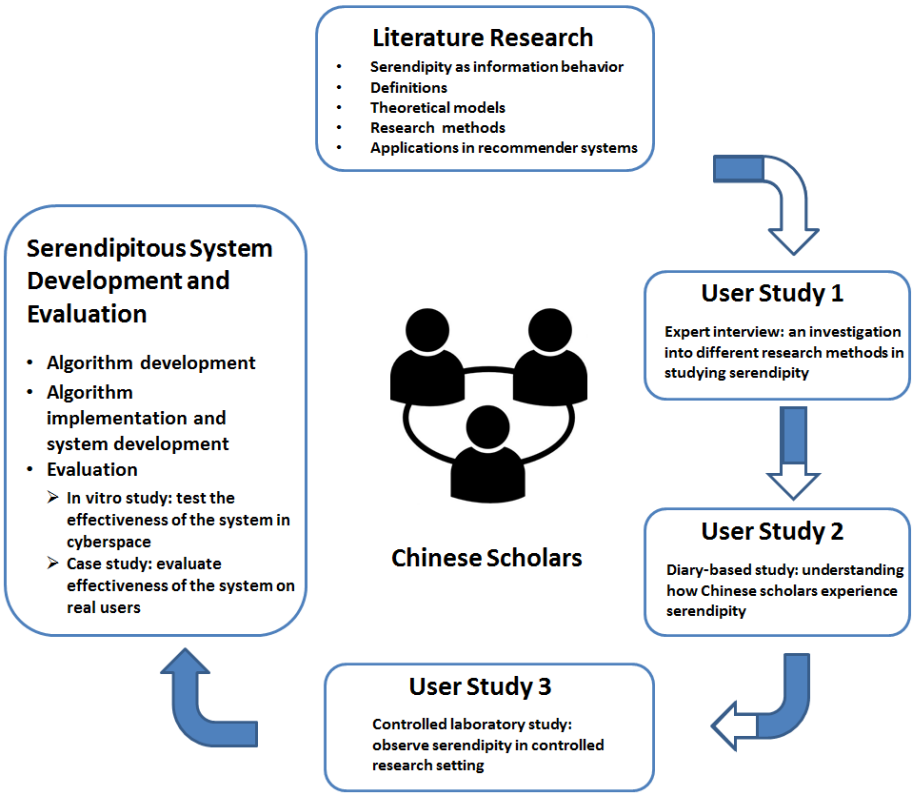


Figure 3-2 Research methods applied in this thesis

3.3.1 Methods applied in Chapter 4 — User study 1: expert interview

The first conducted empirical user study is introduced in Chapter 4, which is an expert interview. Based on the conducted systematic literature review on the employed research methods in serendipity study, the expert interview aims to investigate the following two research questions:

- What are the pros and cons of employing different research methods in current serendipity study in the context of information research?
- What are the challenges confronted when performing these serendipity studies?

Nine experts with rich research experience on serendipity study in information science were invited to the interview. They shared their opinions on the advantages

and disadvantages of employing different methods, and discussed the challenges and existing problems when performing serendipity studies. Their experiences greatly helped the author with a better perception of the methodological issues in conducting related studies, and provided theoretical support to the following empirical studies of this PhD project.

3.3.2 Methods applied in Chapter 5 —User study 2: diary-based study

Chapter 5 introduces a diary-based study on Chinese scholars. This study mainly focuses on addressing the research questions of:

- How Chinese scholars experience serendipity?
- What are the contextual factors that affect Chines scholar's experiencing of serendipity?

While existing literatures all recruit Westerners as participants, this study recruited 16 Chinese scholars as participants and performed a two-week diary study. 64 serendipitous cases were collected within the two weeks, followed by a post-interview to each participant. With the collected data from the Chinese scholars, the study analysed how Chinese scholars experience serendipity and constructed a new context model, which unpack those contextual factors discussed in previous study (Sun et al., 2011) during the processes of serendipity.

3.3.3 Methods applied in Chapter 6 — User study 3: observe serendipity in controlled laboratory setting

When the contextual factors were unpacked in the process of serendipity from user study 2, an interesting finding is that emotion played a role in impacting the experience of serendipity. Existing researchers mainly consider emotion as a positive outcome result from serendipitous encountering (e.g. Rubin et al., 2011; Sun et al., 2011; Makri and Blandford, 2012a; McCay-Peet and Toms, 2015), few of them have discussed whether emotion can inversely to influence an individual's serendipitous encountering. Therefore, this chapter designed a controlled laboratory study to address the research question of:

- What is the role of emotion played in experiencing serendipity?

However, to address such a research question, there are two premises need to be taken into considerations (sub-research questions):

(1) How can a participant's emotions be evoked and captured in a research setting?

(2) How should a research environment be designed to facilitate participant's encountering of serendipity?

Based on the perception of serendipity in information research, the research group first developed an information theory-based algorithm in cyberspace, and embedded the algorithm in a self-developed sketch game through a Wizard of Oz approach. 26 Chinese scholars were invited to the study, they were asked to perform tasks on the sketch game and to fill into a questionnaire during the study process to evaluate whether or not they have experienced serendipity. This study also explored new

research methods to investigate serendipity by employing other HCI equipment (i.e. Eye-tracker and physiological sensors). The developed algorithm and the sketch game provide possibility for participants to encounter serendipity in the controlled research setting, while the adaption of Eye-tracker and physiological sensors made it possible to capture the emotion features evoked by interactions of the sketch game.

3.3.4 Methods applied in Chapter 7 — Serendipitous recommender systems development and evaluation

In Chapter 6, a new information theory-based algorithm is developed, and a user study is performed based on it. The results from the study showed evidence of the effectiveness of the algorithm in providing serendipitous recommendations to users; therefore in Chapter 7 the algorithm is further implemented in real movie dataset to address the research question:

- How to design recommendation technology to facilitate Chinese scholar's serendipitous encountering?

Two experimental evaluations are conducted to validate the developed serendipitous algorithm in the movie dataset:

- An in vitro experiment on a benchmark dataset, in which unexpectedness and serendipity are measured through the approaches proposed by Murakami et al. (2007) and Ge et al. (2010).
- A user study was conducted through a Wizard of Oz approach, and a website was built up with results from the developed serendipitous algorithm in the

recommender system. 12 Chinese movie scholars were invited to the study to assess whether they have experienced serendipity on the developed system.

The results from the study demonstrated that with the help of the developed serendipitous algorithm in recommender systems, Chinese scholars' experiencing of serendipity can be facilitated.

3.4 Conclusion

This chapter starts by a systematic literature review on research methods conducted in current serendipity studies in the context of information research. Through the literature review, it can be found that the study of serendipity is gaining an increasing attention in information research; however, there are two concerns revealed from our review and future studies should pay more attentions on these two parts: one is that exist research methods largely depend on conventional ones (e.g. interview, survey), which relies on participants' memories on serendipitous episodes, and lack of relevant studies to capture instant occurring of serendipity. The other concern is lack of research methods to collect objective data when serendipity occurs.

The chapter then introduces different research methods applied throughout this thesis. By employing these research methods in each chapter, the related research questions are also discussed.

CHAPTER 4

USER STUDY1 — EXPERT INTERVIEW: AN INVESTIGATION INTO DIFFERENT RESEARCH METHODS IN STUDYING SERENIDPITY

Research questions addressed in this chapter:

What research methods can be employed to design studies for serendipity?

[1] What are the pros and cons of employing different research methods in studying serendipity?

[2] What are the challenges when performing serendipitous studies?

4.1 Introduction

In Chapter 3, a systematic literature review is conducted and a general picture of exist research methods employed in current serendipity studies in the context of information research is obtained. 24 empirical studies were screened from the review, and a number of research methodologies were identified. However, due to the slippery nature of serendipity, it is always a challenge to perform empirical studies on serendipity (Makri et al., 2015; Foster and Ellis, 2014). Although the identified literatures successfully employed relevant research methods in empirical studies, the

drawbacks of these methods cannot be identified from the publications. A deep understanding about the pros and cons of these research methods can help us better design the following empirical studies on Chinese scholars. Therefore, an expert interview is conducted, which focused on the methodological parts in performing serendipity studies.

This chapter mainly introduces the conducted expert interview in Section 4.2, and with the systematic literature review introduced in Chapter 3, Section 4.3 proposes the discussions and conclusion on the methodological issues in serendipity studies.

4.2 Expert Interview

4.2.1 Participant

Nine experts with a research experience on serendipity of at least two years were interviewed. These participants were recruited from the SEADE workshop, which was organised under the CHIIR 2016 conference, where a group of researchers were gathered to discuss ongoing research on serendipity. The interviewed participants were seven females and two males, with experience in serendipity study ranging from two year to twenty years (see Table 4-1).

Table 4-1Participant information

Expert	Sex	Research Years on Serendipity
P1	Female	5
P2	Female	7
P3	Female	7
P4	Female	6
P5	Male	6
P6	Female	20
P7	Female	2
P8	Female	2
P9	Female	10

4.2.2 Method

A structured interview was conducted among the nine experts. The interview lasted around half an hour, and mainly focused on the methodological questions when performing serendipity studies. All the interviews were voice recorded with the permission of the participants and were transcribed into Nvivo. See Appendix 1 for the detailed interview protocol.

4.2.3 Data analysis

The collected data were qualitative in nature. A Thematic Analysis (Braun and Clarke, 2006) was conducted to analysis the interview data. Firstly, based on the questions on the interview protocol, a top-bottom thematic analysis was conducted, and a variety of topics like challenges and problems by conventional methods, scenario-related questions, used research methods, advantages and disadvantages of

the method, etc. were coded into different themes. Thereafter, a bottom-up coding process was conducted to further categorise those similar sub-themes to a new theme. All the data coding process was performed through Nvivo 11.

4.2.4 Findings

Table 4-2 lists the research methods employed by these interviewed experts. It can be seen from the table that most of the research methods discussed in Chapter 3 are used by these researchers.

Table 4-2 Employed research methods by different experts

Expert	Used Method for Serendipity Study
P1	Interview, diary-based study, Wizards of Oz, focus group
P2	Interview, online survey
P3	Web-data scanning, interview, online survey, stories
P4	Diary-based study, Wizard of Oz, interview
P5	Interview, focus group, think-aloud
P6	Interview, survey
P7	Survey
P8	Interview
P9	Interview, survey, think-aloud

4.2.4.1 Advantages and disadvantages of different research methods in study serendipity

Table 4-3 lists the reported advantages and disadvantages of the employed research methods in studying serendipity by the interviewed experts. The table is organised with two parts, the first part of data is originally extracted from the

interview study, where the participants discussed the pros and cons of employing these methods into their studies; the second part of the data (italic part in table 4) is a complement of these methods by adopting the discussion from the recent book "Research Methods in Human Computer Interaction"(Lazar, Feng, & Hochheiser, 2017).

Interview

Interview is the most widely used method in understanding serendipity. The advantage of employing interview to study is that it can provide a rich and detailed data from these participants, and based on the provided data, researchers can identify the context details of participant's serendipitous encountering through a retrospective way, and can have a deep look into their thoughts how they understand serendipity when it happens.

> *"Interview can provide quite a lot of information from people in terms of the context or those interactions or those experiences, even though you got the context retrospectively." (P1)*

> *"Interviews and expert reviews can get the qualitative data, collecting information of perceptions that can be really useful" (P2)*

> *"Interview is also very good to get rich details about the experience." (P5)*

However, the disadvantage of interview is that the collected data is based on participant's memories, and it is the retrospective data instead of the instant data when serendipity happens. In addition, during the interview session, participants' respond is based on their sudden opinions and they may probably add more details to the original data in order to play a good role of citizen, and these added details may not be the real situation and may probably bring biases to researcher.

"Sometimes people will add details that don't fit with what actually happened, because they don't remember properly, maybe they just trying to be good citizens and useful participants by giving us details when actually they don't remember the detail." (P5)

"You don't get the information when it happens, so there's a bit of time delay from the finial participants."(P2)

"It relies on people's sudden opinions." (P3)

Further, conducting an interview is never an easy thing; it is often a challenge for a researcher to manage the potential unbound discussions from the participants. Also, the data analysis for an interview usually takes a long time, as Bryman (2001) suggested that one hour of tape usually requires five to six hours to transcribe. In particular, as the definition of serendipity still fails to achieve a defaulted one thus far, different participants are with different perceptions on this phenomenon, which often lead to a perception gap between the researcher and the participants, and this lead to a challenge for a researcher to interpret participants' responses to judge whether an episode can be considered as serendipity or not.

"I found one big challenging is interpreting what [participants] are saying, and a lot of times when I was doing research, [participants] don't talk a lot like us in the serendipity research, they talked like 'I have accident find' or 'I was searching a thing' or 'I saw it on the internet', so it's like picking out what they are actually saying and interpreting that as serendipity." (P8)

Focus Group

Though not identified from the review of the employed research methods in serendipity, three interviewed experts also referred they used focus group to conduct their study. Comparing to interview, focus group can collect data from a group of participants at the same time, and it is flexible for researchers to set different research settings to address relative research concerns. One expert argued that he can collect

excellent feedback from focus group regarding the design strategies of serendipity. The group discussion can support interactivity, and such dynamic situation can encourage participants to raise ideas that probably cannot be identified from one-to-one interview. Researchers can also collect interaction data during the study process and the different layers of data by distinguishing different participant groups.

"It gives you different layer or richness or information when it is a focus group." (P1)

"You can have multiple interactions between different participants, so you can find people or distinct people, and elicit information that cannot be obtained from single individual." (P1)

"Focus groups are good for design, they are good for reaching concerns or design decisions you'd like to make, so they are quite good for me to design digital tools." (P5)

There are also disadvantages of employing focus group. One individual's opinion can often have influence on others. For example, one participant's understanding on serendipity may probably impact the rest participants in the group. The social interaction process may cause pressure to participants.

"Participants might appear pressure due to the social interaction and it then influences their responses in the focus group." (P1)

"Sometimes individual opinions can affect the opinions of others." (P5)

There are also some other challenges when employing this research method. Lazar, Feng & Hochheiser (2017) also point out that questions designed in a focus group are often limited, and it is a challenge for a researcher to select appropriate participants into a group (e.g. research background, familiarity). The group dynamic situation may also lead to some problems, such as how to interpret silence, how to address silent participants, etc. It is a requirement for every researcher with the ability to manage personal conflicts.

Survey

Survey is also widely used in conducting serendipity-related study. The experts used survey mainly on two purposes: the first is to test the data from qualitative studies such as interview. The related questions can be designed, such as those dimensions of testing serendipity in digital environment identified in McCay-Peet and Toms's (2015) study. Another use of survey is mainly to evaluate serendipity, such as by employing Makri and Blandford's (2012b) framework. The survey questions can be easily distributed to participants, either through online methods (e.g. emails, survey websites) or by paper. A huge advantage of applying survey is that it can collect large number of samples with a relatively low cost.

However, the interviewed experts also explained that it is not so easy to enrol enough participants to the survey study, and comparing to other qualitative studies, such as interview, survey cannot have a look into participants' real thoughts on the serendipity episode, and often some designed questions may cause biases to participants (e.g. to ask participant's mood in previous encountering).

> *"It's hard with surveys to get enough people to fill out questionnaires and also it is difficult to seriously look into their thoughts especially when you are doing web-based surveys" (P 2)*

> *"Sometimes only use survey questions would be easy to introduce bias to the data." (P7)*

Diary-based Study

One expert argued she was "a big fan" of diary-based method to study serendipity; this is because diary study is performed in a naturalistic research setting, and it is an engaging method, as the notion of serendipity is often novelty to participants and they are also interested to attempt the diary method to record their encountering on

the mobile phones. A diary study often lasts a certain period of time, hence it is able to record occurred serendipity over time (e.g. time, mood, responses in a serendipitous episode), which can also address the problem of personal infrequency of encountering serendipity. The diary method is often combined with a post-interview in serendipity studies, and thus the provided information of the incidents can help participants to better recall the context when serendipity happens. In addition, serendipity is a user-defined notion, and the diary method can help researchers to define the data that are not easy to be extracted by other methods, for example, by identifying the frequency of serendipitous encountering, the research are possible to distinguish the user group from super-encounterers to none-encounterers (Erdelez, 1997).

The drawbacks of diary method are it is time-consuming and are tedious to capture the data. In particular, in serendipity studies, since diary method is often associated with a post-interview to understand participant's serendipitous encountering during different episodes, it often leads to a relatively long time of the interview. It is also difficult for diary study to get large participant samples; and as it is a naturalistic research setting, participants often encounter serendipity in various situations, and these situations were usually unusual in personal life, it makes the data analysis difficult to generalise research findings such as context, some participants may unsure if they would associate themselves with the findings.

> *"[Diary study] is difficult to get large participant samples, and generalise context was difficult as well, as the context people came up with were really varied and often they were quite unusual in their personal life and things like this, so it was quite difficult to get generalizable findings like laboratory type of context . Some were not sure they would particularly associate themselves for the examples that we get." (P4)*

Lazar et al. (2017) also discussed the disadvantages of dairy method, "participants are sometimes not introspective and not aware of the specifics of what they are doing; they may therefore have trouble recording it in a diary entry" and it is "hard to strike a balance between a frequent-enough series of diary entries and infringement on daily activities (user participation may then trail off)" (p.141).

Think-aloud

Think-aloud is often used in usability test in HCI research. The interview experts also reported that they used think-aloud, often in controlled laboratory study, to better understand participants' experience of serendipity. Using think-aloud in the study can help participants to remember certain cues appeared during the study, and researchers can also understand participant's ongoing cognitive process. An especially benefit of this method in studying serendipity is that the collected data can greatly help participants to memorise the context during the occurred episode when conduct post-interviews, and researchers are possible to collect more detailed and accurate data during the interview session.

> *"Especially in digital environment, I find it is very useful to think-aloud approach, because it was not that exactly asking [participants] specific serendipity experiment in the real moment, but it helped them to remember certain cues and things in the environment, because when I started interview, they had different memories of serendipitous experience online but they have to remember and exactly what caused the serendipity sparks. Once I ask them to go back to do sites, they can easily point out to me certain parts of triggers and even though content is changing." (P9)*

However, it should also be noted that the more use of think-aloud session during a study, it is more possible to interrupt participant's cognitive flow, and will often lead to a longer time to finish the designed task. It is also a challenge for a research to use this method when some participants are not feel comfortable to speak out aloud,

especially when the participant was alone (comparing to pairs, see Hourcade, 2008; Als et al., 2005).

Wizard of Oz

One expert employed Wizard of Oz in studying serendipity. A Wizard of Oz approach is often low time and cost, and human plays the role of a system to provide related information to participants. In most occasions participants are not aware of their achieved information is designed in purpose, but consider the information as a nature feedback from the system. Therefore, this method is helpful to determine conceptual concepts such as serendipity before a real system or when the technology doesn't exist. For example, the interviewed expert developed a prototype of a notebook tool, and based on this tool they employed Wizard of Oz by sending push text suggestions to their participants according to these participants' experiences and interests, which were collected in a pre-interview. Participant's responses to particular kind of information were collected and sometimes including serendipitous information. As a result, it is possible for participants to provide data of their prompt responses when serendipity happens.

The limitation of this method is mainly from the errors made by researchers. It is possible that the role of "wizard", which is played by human, would make errors or mistakes when listening to dictations or typing the words. The expert explained that the original aim of adapting Wizard of Oz in her study was to collect data of participant's prompt responses, but it is also often the case that the participants cannot reach the information sent by researcher timely, which makes the researcher fails to capture the full storyline and context surrounds the accident.

Table 4-3 Advantages and disadvantages of different research methods

Employed method	Advantages	Disadvantages
Interview	1. Rich and detailed data 2. Providing information of context 3. A deep understanding on participant's perception	1. Rely on retrospective data 2. Not real detail data 3. Rely on people's sudden opinions 4. Up to interpretation 5. Take long time to transcribe the data 6. Difficult to manage potentially unbounded discussion*
Survey	1. Easy to be distributed to participants 2. Can test data/finding from qualitative studies 3. Collect large samples of data with a relatively low cost	1. Hard to enrol enough people 2. Hard to understand participant's thoughts 3. May lead to biased data
Wizard of Oz	1. Provide prompt responses when experiencing serendipity 2. Low time and cost 3. Help to determine concepts before the real system or when the technology doesn't exist.	1. Participant cannot reach information timely 2. Human might make errors
Focus group	1. Good for design research concerns 2. Collect interaction data among different individuals 3. Collect different layer of data 4. Group discussion interactively and dynamic research situation	1. Individual opinion impact others 2. Social interaction may produce pressure 3. Limited to a relatively small number of questions 5. Limitations of group dynamics (e.g. difficult to interpret silence, hard to address silent participants) 6. Challenge to manage personal conflicts 7. Challenge to group different participants
Think-aloud	1. Help to remember certain cues 2. Help to memorise the context during interview 3. Good for researchers to understand participant's cognitive process during the study	1. Interrupt cognitive flow during the research process 2. Take longer time to finish a task 3. Some participants may not feel comfortable to speak out aloud.
Web Scanning	Naturally occurring stories	Lack detailed data
Diary-based Study	1. Good to record serendipity happened over time 3. Address the infrequency of encountering serendipity 4.Naturaisc research setting 5. Engaging method 6. Provide information of incidents are helpful to recall in the interview 8. Good for collecting user-defined data (e.g., identifying user group from super-encounterer to non-encounterer)	1. Tedious to capture the data 2. Time-consuming 3. Often lead to long time interview 4. Difficult to get large participant samples 5. Difficult to generalise context 6. Participants are sometimes not introspective and not aware of the specifics of what they are doing; they may therefore have trouble recording it in a diary entry 7. Hard to strike a balance between a frequent-enough series of diary entries and infringement on daily activities (user participation may then trail off)

4.2.4.2 Methodological Challenges in studying serendipity

- Time

Serendipity is a rare phenomenon that one can never pre-determinate when it happens, even during designed studies, and this leads to the challenge of the time for data collection. Conventional research methods such as interview and survey are mainly dependent on a retrospection of a special episode, and there is a lack of investigation into the instance when serendipity occurs.

> *Utilising such [conventional research] methods is that you might get different data when you retrospectively trying to elicit the data, and you might get again different data if you reported as it happens. (P 1)*

> *There is a time dimension for normal research methods, which would not allow us to get the time, although we intended to do that. (P 6)*

- Definition on serendipity

Several participants discussed about the definition of serendipity are a large obstacle to carry out studies, and till now there is a lack of consensus in the research area to a defaulted definition of serendipity. This mainly leads to two methodological related problems:

1) There may be a perception gap between participants' responses and researcher's identification. Participant normally won't directly report their encountering as serendipity, and they may describe this phenomenon by other similar words, such as by accident, by chance, or even without an obvious description. The researcher need to identify the notion of "serendipity" based on various responses from different participants, and this interpretation process inevitably results to a potential challenge for researchers not to bias participants' original understandings:

It's like picking out what [participants] are actually saying and interoperating that as serendipity, that's sort of a challenge. My initiate saying of this is what's happening and you are encountering serendipity, but that may not what [participants] are actually experienced. (P 9)

The problem is that you are also the same time by sense of people's respondents, and it is difficult not to bias people's responses because you have to define the notion of serendipity and to ensure back and assistance to response. (P5)

2) It poses a challenge to design and prepare a study. Lacking a census definition on serendipity makes every individual has his/her own definition on this special term, and it is not easy to design survey or interview questions that target on individual's conceptual perception on serendipity. For example, during this interview study, a scenario is provided to all these experts (see question 3 in appendix A). The original use of this scenario was aimed to investigate what methodologies these experts would adapt if they were under the condition of conducting the study of this scenario, but it was unexpected that over half of the experts (five out of nine) disagreed that this scenario was a case of serendipity. This part of data was therefore eliminated during this report and it led to the re-thinking of how to choose a more appropriate scenario in future studies.

- Individual differences

The participants also reported individual differences as factors that impact the study of serendipity. The differences mainly cover the following two folds: first, the ability to memory is individually different. For example, one expert exemplified that during their conducted interviews, some participants can recall any experience of serendipity quite well, while some participants cannot remember the details of the encountering. Another challenge is the experiencing of serendipity. As identified by

Erdelez (1997), people encounter serendipity can be divided into five categories, ranging from super-encounterers to non-encounterers. One expert (P5) argued that she has a colleague who claimed that she would never experience serendipity, while the expert considered herself frequently encounter serendipity in her daily life. The individual differences make researchers difficult to manipulate the recruitment of participants to perform the study.

4.3 Discussion and Conclusion

Based on the literature review in the previous chapter and the conducted expert interview, it can be found that numerous HCI research methods have been employed in studying serendipity. While continuous processes are achieved to uncover the veil of serendipity in recent decades, this paper provides methodological contribution to serendipity studies in the context of information research.

Foster and Ellis (2014) argues "there must be a way for the researcher to learn when a serendipitous recognition occurs, either at the moment of the event or at some later point" (p.19). While the study has found that most of current empirical studies have successfully employed methods like interview, survey, focus groups, etc. to understand the "later point" after serendipity by retrospect into participants' memories, few studies so far can catch the "moment of event" when serendipity occurs. The identified controlled laboratory studies demonstrated possibilities to observe serendipity under controlled research setting, and such lab-based studies may be a solution to capture more detailed data during the moment when serendipity occurs if they are carefully designed. For example, Erdelze (2004) proposed using eye tracking technology to capture subtle changed in user's attention shifting; and Miwa et al.

(2011) successfully captured participant's eye tracking data and provided strong evidence of their results that "information encountering sometimes led to a temporary deviation from the initial task, but may have resulted in a change of the topic of the paper and/or destination of the trip". Except eye-tracking, it is also possible to employ other HCI method by designing lab-based studies to better capture the moment of serendipity, such as muscular and skeletal position sensing, motion tracking , and use physiological tools to collect the physiological data (e.g. electrodermal activity, cardiovascular signals, respiration, brain activity, muscle tension), as introduced by Lie et al. (2017). I believe by a deliberation of the possible disadvantages of controlled laboratory study and through a careful design, it is possible for researcher to capture data at the moment when serendipity happens.

From van Andel's (1994) description of serendipity as "the art of making an unsought finding", to Cunha's (2010) definition of "the accidental discovery of something that, post hoc, turns out to be valuable", and more recently the definition given by McCay-Peet and Toms (2015) as "an unexpected experience prompted by an individual's valuable interaction with ideas, information, objects, or phenomena", "what happens when we, in unplanned ways, encounter resources (information, things, people, etc.) that we find interesting" by Björneborn's (2017); it seems there is still a long way to reach a consensus on the definition of serendipity among these researchers. However, despite a consistent definition, the frameworks can be extended to identify serendipity. For example, Foster and Ford's (2003) classified four different categories of serendipity, known as "reinforcing or strengthening the researcher's existing problem conception or solution", "taking the researcher in a new direction, in which the problem conception or solution is re-configured in some way" (p.330), "the

unexpected finding of information the existence and/or location of which was unexpected, rather than the value", and "the unexpected finding of information that also provide to be of unexpected value (a) by looking in likely sources; (b) by chance"(p.332). Based on a ground theory coding process, Makri and Blandford (2012b) proposed their own classification framework of serendipity; by considering three elements of "unexpectedness", "insight" and "value", they successfully identified their serendipitous cases, and this framework is also adapted in recent studies (e.g. Points et al., 2015; Kefalidou and Sharples, 2016) to identify whether or not participants have experience serendipity. This framework is further extended in a recent study (Zhou, Sun, Wang & Sharples, 2018), and through an identification of the contextual factors (i.e. external context, social context and internal factors) from a diary-based empirical study, the "unexpectedness" participants encountered during a serendipity encountering can be further categorised into "unforeseen means and/or unexpected content of information", and then "making connections with a visceral need, conscious need or previous knowledge/experience", and finally lead to a "substantial and/or emotional value". Rather than to give a definition on serendipity, these proposed frameworks provide another solution for researchers to identify serendipity among a variety of collected examples.

While a majority of discussed research methods collect data mainly depend on a small sample of participants (e.g. interview, focus group, diary studies, controlled laboratory study), it is inevitable that there will be individual differences in the frequency of encountering and perception on the notion of serendipity. A possible solution is to employ online crowdsourced studies. According to Lazar et al. (2017), "Crowdsourcing studies use online platforms to collect data from participants over the

web, usually through the use of web software designed to enrol participants, provide training, and complete relevant tasks" (p.429). This kind of study allows researchers to collect data from large samples, and thus is possible to generate more statistical conclusions comparing to a small sample of participants. The selective blog mining or what the expert called as "web scanning", which was dependent on analysing user generated contents online (e.g. blogs and posters), is a typical crowdsourced study. Future studies can pay more attention on this part to generate stronger research findings.

An implication from the interview study is to use multiple complementary methods to conduct serendipity research. Having discussed the pros and cons of these research methods employed in serendipity studies, it is possible for researchers to better design the methodology issues according to different research purpose. For example, interview is a good solution to look into participant's deep thoughts in a serendipitous episode, but it depends on participant's memories. A combination of diary study to collect the instant episode or combing the controlled laboratory study with recorded pictures, videos or log files can better help participant's memorise of the context when serendipity occurred. In addition, a think-aloud session, especially conducted in the controlled laboratory studies, would bring much benefit for researchers to catch up participants' cognitive processes, and also provide possible clues for the post-interview or post-surveys after the study. While there are individual differences in experiencing serendipity, apart from crowdsourced studies, a prior survey can also be tested to distinct different group of participants from super-ecnounterer to non-encoutnerer, and it sometimes can be helpful to conduct diary-

based studies or lab-based studies by avoiding choosing those non-encoutnerers as participants.

As argued by one of the interviewed expert, “there is a real scope in developing new innovative, original research methods for serendipity, and I feel the existing methods help us in some ways, but they also constrained us.” Through the discussion on existing research methodologies of serendipity, future studies can pay more attentions to these issues and conduct related studies more effectively and efficiently.

CHAPTER 5

USER STUDY2 — DIARY-BASED STUDY: UNDERSTANDING HOW CHINESE SCHOLARS EXPERIENCE SERENDIPITY

Research questions addressed in this chapter:

What is the role of context played in Chinese scholars' experiencing of serendipity?

[1] How Chinese scholars experience serendipity?

[2] What are the contextual factors that affect Chinese scholars' serendipitous encountering?

5.1 Introduction

The literature review chapter (Chapter 2) has showed that very few of existing studies have examined the occurrence of serendipity from a systematic perspective of context. The empirical studies by Points et al. (2015) showed contextual factors such as location, activity and focus can influence a user's experience of serendipity. Kefalidou and Sharples's (2016) study also found that time, location, and the content of a text message can also impact a user's experience of serendipity. Serendipity, as part of a wider behaviour model, is considered as "the product of context" (Foster and

Ellis, 2014, p.18), and the role of context in fostering serendipity deserves to be understood in its own right.

The only related study on the role of context is conducted by Sun et al. (2011), where Schmidt's (2000) context model is adapted and the relationship between the role of the individual and their context in serendipity is denoted by considering different elements, such as an individual's level of attention, pressure, and focus under the effects of the physical environment, the social environment and the time (See Figure 5-1). However, limited by the perception of serendipity at that time, the authors only discussed these elements as different factors having an influence on participants' readiness to experience serendipity, and failed to make a further examination of how these elements would act during the separated processes of serendipitous encountering.

In addition, in the discussion part of Chapter 2, I have pointed out that current research findings on serendipity in the context of information research were mainly based on Westerners as participants, and whether or not these research outcomes (e.g. theoretical models, employed research methods, perceptions on serendipity) are also pervasive on other culture groups such as Chinese are still unknown. As a matter of fact, a number of studies have confirmed that culture has a significant impact on individuals' information behaviours. For example, Yeh (2007) put forward a model of culture and information behaviour, in which he argued that people from different cultures will develop different traditions, prejudices and habits, and these are all significant elements which will impact on an individual's information-seeking behaviour. In addition, certain Culture Neuroscience (CN) studies have demonstrated that individuals with different cultural backgrounds can perform differently in the psychological processes, such as the cognitive processing styles in perception (Ji, Peng, & Nisbett, 2000), memory (Wang & Conway, 2004), attention (Kitayama,

Duffy, Kawamura, & Larsen, 2003), and causal attribution of events (Morris & Peng, 1994). Serendipity, as part of information seeking behaviour (Erdelez, 1997), is also considered to be highly related to psychological processes, such as the "prepared mind" (McCay-Peet & Toms, 2015; Rubin et al., 2011), sagacity (McBirnie, 2008), and attention (Erdelez, 2004; Rubin et al., 2011); therefore, whether or not culture has an influence on experiencing of serendipity deserves to be investigated in its own right.

Based on the above discussion, this chapter performed a new empirical study among a group of Chinese scholars with the following research aims:

- To identify whether the current understanding of serendipity can also be adapted to Chinese scholars;
- To further investigate the role played by context during the different processes of experiencing serendipity;

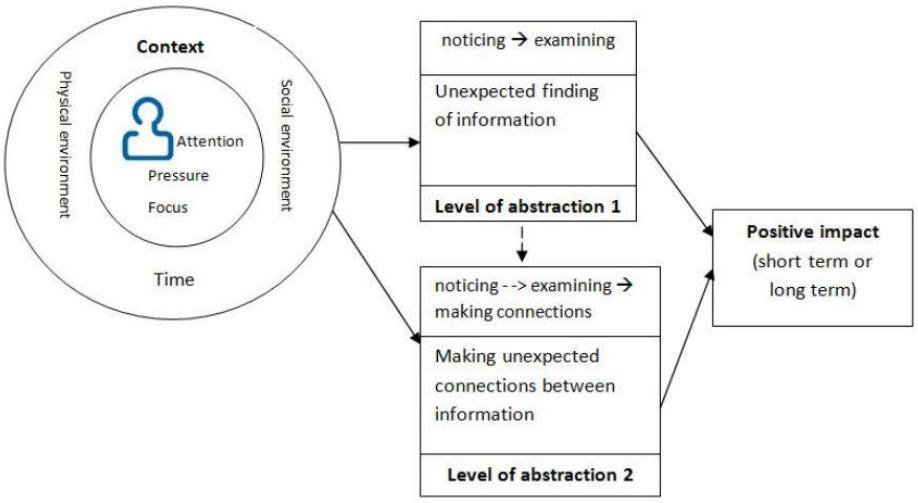


Figure 5-1 The previous context model for experiencing serendipity (Sun et al., 2011)

5.2 Background

5.2.1 Context in serendipity research

In recent decades, a number of researchers have performed different studies with respect to serendipity, although these researchers have not reached a consensus on the definition of serendipity. For example, van Andel (1994) defines serendipity as "the art of making an unsought finding", while Fine and Deegan (1996) give the definition of serendipity as "the unique and contingent mix of insight coupled with chance". More recently, McCay-Peet and Toms (2015) contend that serendipity is "an unexpected experience prompted by an individual's valuable interaction with ideas, information, objects, or phenomena", while the term serendipity is defined in Björneborn's (2017) paper as "what happens when we, in unplanned ways, encounter resources (information, things, people, etc.) that we find interesting". Rather than giving a definition, Makri and Blanford (2012) identified three key elements for serendipitous encountering: unexpectedness, insight and value.

However, regardless of the various definitions, it is well accepted by information researchers that serendipity is an integral part of information behaviour, and "context" is a significant concept when studying information behaviour, as argued by Case and Given (2016):

> *The seeker—whether actively looking for information or receiving information through serendipity—exists in an environment that partially determines, constrains, and supports the types of needs and inquiries that arise. The seeker also has his or her own memories, predispositions, and motivations—an internal environment of influence. (Chapter 3, p. 48) and Context determines much of a person's perceptions throughout the*

[information seeking] process, and it affects one's choice of sources and meanings. (Chapter 11, p. 351)

Björneborn's (2017), who adopted the theory of affordances from Gibson (1979), also argues that serendipity can be viewed as an affordance, which should never reside inside the environment alone nor inside people alone, but should be viewed as the relational phenomenon between people and a given environment. In a similar vein, after a review of serendipity studies in information research, Foster and Ellis (2014) concluded that serendipity does not exist within a vacuum, but is "the product of context" (p.18). Some empirical studies also demonstrate contextual factors affecting an individual's experiencing of serendipity. For example, through a "Wizard of Oz" approach, where users received text messages/suggestions from a group of "wizards" based on users' notes in an app "SerenA", Points et al. (2015) found that those contextual factors such as location, activity and focus can influence a user's experience of serendipity. Similar findings can be found in another paper (Kefalidou and Sharples, 2016), where the contextual factors such as time, location, and the content of the text message can impact a user's experience of serendipity. McCay-Peet and Toms (2015) have found that those environmental factors which are trigger-rich, enabling connections and leading to the unexpected can help users facilitate serendipity in a digital environment. Such ongoing research findings provide substantial evidence that context does play a vital role in people's experience of serendipity.

5.2.2 A further discussion of context

From a review of the existing studies on the issues of context and serendipity, it is evident that none of them have systematically discussed the term "context", nor how it may influence the different processes during a serendipitous encounter. Björneborn (2017) used the term "affordance", McCay-Peet and Toms considered "environmental factors", while Kefalidou and Sharples' (2016) description of context also refers to a user's different activities. Foster and Ellis (2014) argued that "even context is debatable and has been the subject of exploration in its own right" (p. 18). Case and Given (2016) considered context to be "ill defined", but also highlighted its important role when integrated in human information behaviours. Taken together, I believe there is a need to probe this special term "context".

Based on a review by Courtright (2007), the study of "context" in information science has shifted from a "system-centred" to a "user-centred" stance. A previous "system-centred" view regards context as an "objective reality" (Talja, 1997), which has served as a backdrop for those environmental factors or variables that exist objectively around the information actor, and can therefore be enumerated by the researcher. Such a view of context is also labelled as "objectivist" (Talja et al., 1999), which presents context as a set of entities that can be conceptualised independently to influence a participant's information practices (e.g. temporal or spatial conditions, problem situations, etc.). However, taking only those environmental variables into the consideration of context fails to shed light on the variability among actors in the same or similar settings. The information actors can carry out actions independently and differently in response to the variability of the environmental factors in their information practices. Therefore, an increasing number of researchers have now

attempted to examine the role of context from the viewpoint of the information actor. This "user-centred" view emphasises the role of information actors during their information practices, and considers the information activities in relation to the contextual variables and influences. Various models have been constructed to support this kind of view, such as Wilson's (1981) information seeking model where an individual's physiological, affective and cognitive needs are located in the concentric layers at the root of motivation towards the information seeking behaviour. Foster's (2004) nonlinear model for interdisciplinary information-seeking also highlights information seekers' feelings and thoughts, coherence, knowledge and understanding as the internal context to influence information seeking behaviour. Although such a person-in-context stance is being accepted by more and more researchers, there are also critics who argue that these models do not account adequately for the mutual interactions of contextual factors, especially the social interactions. Each individual is conceptualised as a social actor (Lamb et al., 2003) and knowledge as inherently social (Talja, 1997). Therefore, information actors should construct information not only through their physiological or affective needs but also through social interactions. Rather than simply observing the information actors' behaviours or recording their views, the relevant discourse should also be taken into consideration when trying to gain an understanding of the role of context in information research (Given, 2002; Sundin, 2002; Talja et al., 2005). Taken together, Courtright (2007) suggests the combination of multiple methods to paint a comprehensive portrait of context, which should not only try to capture any environmental variables around the information actor, but should also try to understand their mind-sets and follow the links across their multiple social settings.

Following the above discussion, in this chapter I discuss the environmental variables as the "external context", the mind-sets relating to the role played by the information actor as the "internal context", and the social settings around the actor as the "social context".

5.2.3 Existing serendipity models in relating to contextual factors

In Chapter 2, apart from the previous research model (Sun et al., 2011), I also reviewed the other six existing models for serendipity, five of which are process-oriented (i.e. Erdelez, 2004; McCay-Peet and Toms, 2010 & 2015; Lawley and Tompkins, 2008; Makri and Blandford, 2012a), while the sixth is based on the essence of serendipity (Rubin et al., 2011).

On further reflection of these models, I have found that from a "user-centred" point of view, these models all partly refer to the contextual factors, especially those relating to the information actor, as is listed in Table 5-1. Erdelez's model requires users to notice the background problem, and such an "ability to notice" can be considered as part of the internal contextual factors that affect an individual's experiencing of serendipity. In addition, her model is useful for understanding part of the process post-encounter, but it fails to cover what happens beforehand (e.g. whether or not external factors played a role to trigger the encounter). McCay-Peet and Toms's (2010) model identified "precipitating conditions" as "active learning" (internal context) and "social networks" (social context), and as a requirement for a "trigger" (e.g. text, images, audio) to facilitate serendipity. However, as an early model in knowledge work, this model fails to look into how the "precipitating conditions" would impact the process of serendipity. In their updated model (McCay-

Peet and Toms, 2015), they further highlighted how the "trigger" confirms the "noticing" element of the process of serendipity, and identified the three forms of triggers as "verbal" "textual" and "visual". They further proposed different external factors (trigger-richness, highlight triggers, enabling connections and enabling capturing) and internal factors (openness, a prepared mind, the ability to make connections) that may influence the perception of serendipity, but these factors, especially the external factors, are not discussed from the perspective of context. The other two mental-process models focus mainly on the perceptual process required for a serendipitous episode, and also discussed some contextual factors. For example, Lawley and Tompkins (2008) considered a "prepared mind" (internal context) and an "unexpected event" (external context) as necessary components in a serendipitous episode, while similarly, Makri and Blandford (2012a) considered how "unexpected circumstances" (external context) and "insight" (internal context) can lead to making new connections, and they also found that, although not directly reflected in their model, their participants' moods or feelings (internal context) can impact the openness to making connections. Similarly, chance (external context), a prepared mind, an act of noticing and surprise (internal context) can all be considered as contextual factors that are referred in Rubin et al.'s (2011) model. However, a systematic discussion from the perspective of these contextual factors cannot be drawn from these existing studies.

Table 5-1 Contextual related factors in existing serendipity models

	External Context	Internal Context	Social Context
Erdelez (2004)		Notice	
McCay-Peet and Toms (2010)	Trigger	Precipitate condition	Precipitate condition
McCay-Peet and Toms (2015)	Trigger	Openness Prepared mind Ability to make connections	
Lawley and Tompkins (2008)	Unexpected event	Prepared mind	
Makri and Blandford (2012)	Unexpected circumstances	Insight	
Rubin et al. (2011)	chance	Prepared mind act of noticing surprise	

5.3 Research Method

One major concern raised in Sun et al.'s (2011) study is portability. Some of the participants were unwilling to carry around an additional mobile device (which installed their developed diary-application), and thus affected their serendipitous experiences. To overcome this particular drawback, in this study I used the social media platform "Wechat" to replace the diary application. The main reasons for choosing "Wechat" were: 1) it covers similar functions to the diary application, and different types of data can be recorded and transferred (i.e. text, video, audio, and image); 2) participants had no concerns about portability problems, and no additional package needed to be installed on participants' own mobile phones, as they were all frequent users of "Wechat", and were quite familiar with its functions; 3) "Wechat" is

a social media platform, so it also has the function of allowing direct communication between participants and the researchers, and if participants had any problems during the experiment period, they were able to send messages to the researchers and receive immediate responses; 4) it had the advantage of allowing the researchers to send a "reminder" to participants each day, to help to make them aware they were in an experiment situation.

5.3.1 Participants

16 Chinese PhD students (eight males and eight females) were recruited to take part in this follow-up study, with each participant having had at least 12 months' research experience. The reason for choosing PhD students as participants were mainly because: 1) following the research findings of Foster and Ford (2003) which showed that serendipity is experienced widely among researchers, and PhD students are a group of scholars dedicated to research projects who are easy to access; 2) Sun, et al.'s (2011) previous study recruited 11 PhD students and received 23 serendipity cases within a week. This implicated that it was a feasible solution to recruit PhD students with which to conduct such a diary-study. Detailed information about all participants is listed in Table 5-2. All the participants' names reported in this study are aliases.

Table 5-2 Participant Information

No.	Research Interests	Gender	Research Time (Months)
1	History	Male	36
2	Mechanical Engineering	Male	12
3	Computer Science	Male	24
4	Civil Engineering	Male	24
5	Environment and Energy	Male	18
6	Exhibition Design	Male	72
7	Computer Science	Male	13
8	Fluid Mechanics	Male	36
9	Operation Management	Female	20
10	Chemical Engineering	Female	13
11	Consumer Behaviour	Female	16
12	Pedagogy	Female	12
13	International Economics	Female	38
14	Pedagogy	Female	28
15	Environment and Energy	Female	36
16	Chemical Engineering	Female	18

5.3.2 Procedure

1) Pilot Study. A pilot study was performed with two participants (one male and one female) at the University of Nottingham Ningbo China for a period of four days. The detailed experiment issues (e.g. time arrangement, interview preparation) were all determined according to the pilot study. Four serendipitous cases were collected from the pilot study.

2) Pre-interview. Each participant was invited to a short interview (around 30 minutes) before the empirical study. The research purpose was introduced, and participants were invited to collect any cases they considered as serendipity during a period of two weeks, either on the Web or as part of their daily activities (e.g. reading,

research, and socialisation). In addition to introducing the research purpose to each participant, I also conducted two additional operations during the pre-interview, as follows:

(a) Each participant's initial understandings of serendipity were collected. During the interview process I found that each participant reported that this was the first time s/he had heard about the concept of serendipity. To better support the study, I then carefully introduced participants to this concept. First, by presenting the definition of serendipity from the Oxford Concise English Dictionary: "*the occurrence and development of events by chance in a happy or beneficial way*". I then provided participants with the following example from the pilot study:

> *I was trying hard to download a journal paper which could be very relevant to my research. However, when I checked in our university databases, it was not available to download and payment was required to get access to the paper. Then, one day when I was searching for other research papers, a web link of the paper turned up on the screen. Being curious, I clicked the link and it asked me to register in a platform called Research Gate. I followed the registration and was then amazed to find that the author of the paper was also a member of Research Gate, so I followed him on Research Gate and sent him a request for a copy of the paper, and he sent me a copy of his working paper free of charge! It was really exciting for me to get the paper by such a chance! (Pilot study Case 4)*

Participants were highlighted that this example was simply provided to help them to understand the concept of serendipity, and by no means to restrict them to a particular type, stressing that there are various examples. They were instructed that if they experienced any encounters which they considered as serendipity, they should send the researchers a relevant message.

(b) Participants were introduced to the group created using the social media tool "Wechat". To achieve better research results, a specific interface was designed and

instructed to participants about its functions (Figure 5-2). A detailed description of the interface is provided in Appendix 3.

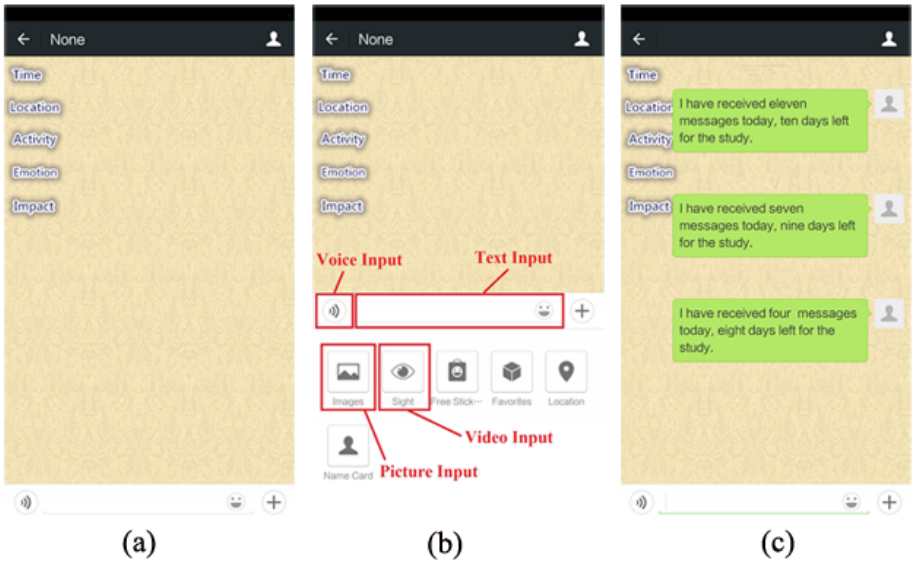


Figure 5-2 Wechat as research platform: (a) designed interface; (b) different input sections; (c) daily reminder sent to participants.

3) Two-week study. The experiment lasted two weeks. Participants were required to use the tool provided to record their serendipitous experiences, and return them to the researchers within two weeks. All the sent data was only visible to the researchers. In addition, at approximately 10:30pm each day, a reminder message was sent to each participant by the researchers to better provide them with a research context (Figure 5-2-c). The time chosen for sending the reminder was based on the pilot study and observation of the routines of most participants.

4) Post-interview. Each participant was invited to a post-interview at the end of the study. The interview was conducted within one week and lasted for approximately one hour with each participant. It was semi-structured and qualitative in nature and centred on participants' recorded serendipitous encounters, as well as participants' experiences of the research method.

5.3.3 Data collection

Two types of data were collected: the recorded diary data of the participants' serendipitous experiences and the post-interview data. A total of 62 serendipitous records were received, each describing a case which the participant regarded as serendipity. The records for each participant were printed out to help them reflect on their experiences during the post-interview. All interviews were recorded and transcribed by the interviewer.

5.3.4 Data analysis

The collected data are qualitative in nature. A Thematic Analysis (Braun and Clarke, 2006) was conducted to code the interview data. First a top-bottom theoretical thematic analysis was conducted to investigate how the participants experienced serendipity. This part of coding was started by identifying the themes drawn from previous study (Sun et al., 2011), where the nature of serendipity was identified according to two different levels of abstraction, and the value of serendipity. The first level identified the "unexpected finding of information" by considering different combinations of three components: whether the information was directly related to the activity being undertaken by the individual (non-activity-based vs. activity-based); whether or not the information encountered was unexpectedly valuable to the encounterers (unexpectedly valuable or not); and whether the information was from an unexpected or likely source. The second level identified the making of unexpected connections between different pieces of information, people and ideas.

A bottom-top inductive thematic analysis was then used to identify any contextual factors which existed in the serendipity cases. Initially, I identified a number of categories, including the time for experiencing serendipity (i.e. a.m., p.m., and across

time periods), the different locations in which serendipity occurred (e.g. office, dormitory, classroom, , library, etc.), the different activities during which serendipity was experienced (e.g. travelling, surfing the Internet, attending seminars, talking to classmates, talking to friends, etc.), and a category more related to an individual's cognitive or psychological characteristics, such as memories, an amount of thinking, expertise, previous needs, instantly raised needs, and emotions (see Figure 5-3, for examples of the coding for the pilot study case). After a comparison of the categories, those with overlapping meanings were grouped into possible themes. As a result, this layer of analysis was concluded into three major themes of: external context (i.e. time, location, and personal status), social context (e.g. different social counterparts) and internal context (i.e. precipitating conditions, sagacity/perceptiveness and emotions). It should be noted that the precipitating conditions include visceral needs, conscious needs and previous experience/knowledge.

After finished coding the themes of the contexts, especially the internal context themes of the precipitating conditions, I carried out a full review and found that the original coding of the first level "unexpected finding of information" could also be considered as a process of making connections between the encountering and the precipitating conditions. As a result, this part of the framework was re-coded into three different themes of unexpectedness, connection-making and value, which were further expanded into the sub-themes of "unforeseen means of encountering information", "unexpected content of the encountered information" and "both". The theme of "connection-making" was further expanded into "connection-making between unexpectedness and visceral needs", "connection-making between unexpectedness and conscious needs" and "connection-making between

unexpectedness and previous experience/knowledge", while the theme of "value" was expanded into the sub-themes of "substantial value" and "emotional value". This will be discussed in detail in the following sections.

Interview Data Extract: Pilot study P1	Code For
I was trying hard to download a journal paper which could be very relevant to my research[a]. However, when I checked in our university databases, it was not available to download and payment was required to get access to the paper. Then, one day when I was searching for other research papers[b], a web link of the paper[b] turned up on the screen[1]. Being curious[c], I clicked the link and it asked me to register in a platform called Research Gate. I followed the registration and was then amazed to find[d] that the author of the paper was also a member of Research Gate[2], so I followed him on Research Gate and sent him a request for a copy of the paper, and he sent me a copy of his working paper free of charge! It was really exciting for me to get the paper[e] by such a chance[3]! Q: When did you receive the paper from the author? A: The next afternoon[f] when I was working in my office[g].	Framework of serendipity: 1. Unexpected encountering 2. Connection-making 3. Value of the encounter Contextual factors: a. Existing need b. Searching online c, d & e. Emotions related f. Across time period g. In the office

Figure 5-3 Example of the analytical rationale used for the data analysis

5.4 Participants' Perceptions of Serendipity

By analysing the 62 reported serendipitous cases, I found that the Chinese scholars also conformed to the framework of experiencing serendipity according to the three main processes of: encountering unexpectedness, connection-making and finally leading to a valuable outcome.

5.4.1 Unexpectedness

Three different channels were identified from this empirical study to facilitate the likelihood of encountering unexpectedness during the new study:

(i) Any unforeseen means by which a participant encounters a piece of information. An example, which can better explain the identification of this element, is provided in the following case:

> *[In a training session] a student delivered a talk on fire extinguishers several days ago, which made me recognise that I had never noticed there is a fire extinguisher in my lab before, and I raised some concerns, such as what were they used for? How did they work? I had all these concerns resolved today by accidently attending a 'fire alarm introduction' presentation. (Case 7)*

The participant (P3) reported she had learnt about the principles of fire extinguishers from a talk during a training session. She then raised the need to collect relevant information about fire extinguishers around her lab setting (e.g. where they are located in the lab, and how to use them). However, this need was not addressed at that time and the participant forgot to address the need after the talk had finished. As explained by the participant during the interview, "I just thought about it in my mind and didn't write it down during the talk. Actually, when the talk finished, I just forgot about it". It was not until she accidently attended a related "fire alarm introduction" that she realized the need again, and found the answer to this need during the presentation. Actually the answer to the participant's need was not unexpected to her, but the way that she received the answer made her feel it was "unexpected", as attending such a presentation was not her original schedule, "I even didn't know about such a presentation, but one of my friends just asked me to accompany her".

(ii) The content of the encountered information brings unexpectedness. In some other cases, it is the content of the information that leads to a sense of unexpectedness:

My instructor from an academic training session asked his students to present an article during the session which he had just handed out to us. The article was about a wind-up radio which greatly aroused my interest in radio technology. I had never thought I would learn about wind-up radios in this training. (Case 6)

In this case, the participant was situated in a certain context (a training session), and it was the sudden appearance of information (about the radio) which was interesting to him and resulted in his feeling of "unexpectedness".

(iii) Both the unforeseen means and content of the encountered information bring a sense of unexpectedness:

There was a seminar, but I didn't pay attention as it seemed not so relevant to my research. However, I was required by my supervisor to attend. It was difficult for me to concentrate at the beginning of that seminar. My interest was aroused when I noticed from the PPT that the lecturer had used the same simulation software as me and he used a graphical way to present the results in his research, which I had never considered before. I found it was really useful! (Case 1)

This is an interesting case, not only because of the unexpected information gained from the seminar by the participant (i.e. presenting results with graphics), but also the unexpectedness of attending the seminar, as explained by the participant:

In the beginning, it was not my intention to attend the seminar as I thought it was not so relevant to my research. I just attended by accident and it was not something I had planned to do. Furthermore, I didn't expect there would be such useful information which I could take away from the seminar. Therefore, I would consider it as serendipity. (P 1)

From this response, it is evident that both the useful information she received from the seminar, and the way she received the information (by attending the seminar accidentally) functioned in her coming across this serendipitous experience.

5.4.2 Connection-making

Previous study found that connections can be made between different pieces of information, people and ideas (Sun et al., 2011), and it is a level of abstraction that can lead to a positive impact. In this chapter, this process of connection-making was further expanded by identifying the different internal contextual factors of precipitating conditions:

- Connections made between unexpectedness and visceral needs.
- Connections made between unexpectedness and conscious needs.
- Connections made between unexpectedness and previous experience/knowledge.

The term "visceral need" and "conscious need" originates from Taylor's (2015) work, which characterised four different levels of information needs during the interaction between an information seeker and a librarian. An information seeker may begin with an unexpressed need in mind (what Taylor calls a "*visceral need*"), and then such a need becomes "*conscious*" with accumulated information (e.g. by talking to the librarian), and turns to a "*formalised need*" with a qualified and rational statement, which finally leads to a "*comprised need*" that can be presented to information systems. Although Taylor's framework is usually used to describe the negotiation process between an information enquirer and an information specialist, which is not exactly the process of encountering a serendipitous episode, I do find that the characteristics of such "visceral needs" and "conscious needs" conform to the situations reflected in the participants of this empirical study. For example, Taylor argues that a "visceral need [not] existing in the remembered experience of the enquirer, [probably] is inexpressible in linguistic terms, [and can] change in form,

quality, concreteness, and criteria as information is added." In this study, I also found some participants did not raise a need, which was not previously in their memory or experience, until they encountered unexpected information. While a "conscious need", as described by Taylor, is a "within-brain description", it is quite similar to a previously unaddressed concern/question by a participant. This part will be discussed further in Section 5.5.3.1.

5.4.3 Value

Participants pointed out that they would only consider any unexpected experiences as serendipity if they were offered with some benefit. Two types of value arose from the study: *substantial value* and *emotional value*. *Substantial value* refers to a value that brings beneficial results or outcomes to the participant (e.g. finding the answer to a previous concern), while *emotional value* refers to a value caused by an emotion which is aroused when a participant encounters unexpected information (e.g. the emotional satisfaction of recalling previous memories). This part will be discussed further in Section 5.5.3.3.

5.5 The Role of Context

5.5.1 External context

"External context" refers to the ambient conditions surrounding the participant, and the three different external contextual factors which have been identified from the empirical study are time, location and personal status (Table 5-3).

Table 5-3 External Context Factors

External Context Factors	Elements	Numbers of Serendipitous Cases
Time (55 available cases)	A.M.	8
	P.M.	39
	Across different time periods	8
Location (58 available cases)	Office	29
	Seminar room	11
	Dormitory	6
	Other random places	12
Personal Status (34 cases)	Leisure	19
	Seminar	9
	Working/studying	9

(1) Time. There were seven out of 62 collected cases for which the participants were unable to recall the time of the encounter. Among the remaining 55 available cases, only eight happened before noon (a.m.), while the remaining 47 cases happened after noon (p.m.). It is evident that different time periods during the day contributed distinctly to the development of serendipitous experiences. The final eight cases were reported across different time periods, where the participant was engaging with an ongoing activity until s/he recognised the occurrence of serendipity after some time. Existing research has demonstrated that different times of day can impact human performance (Fröberg, 1977), and even the cognitive and evaluative efficiency of individuals (Natale et al., 2003). The participants also reported that they were more engaged in different activities in the afternoon, as a result of which it was also more likely that they would encounter serendipity, as explained by one of the participants:

> *Personally speaking, I find myself more conscious about the concept of serendipity in the afternoon or evening than in the morning, and normally I'm more engaged in the*

afternoon. So I think that's the reason why I always send you messages in the afternoon. (P 4)

(2) Location. Some locations (e.g. libraries) are richer in resources (e.g. books) than other locations (e.g. canteens). Therefore, it is intuitively sound to assume that locations may influence the occurrence of serendipity. 58 of the cases reported by participants indicated where their serendipitous experiences had taken place (the location of the remaining four cases could not be recalled). The data showed that 29 cases happened in an office environment, followed by 11 cases in a seminar room, six cases in a dormitory and other random places (laboratory, café, library, etc.). Following a further look into the office environment, which produced most serendipity cases during the study, I found that there were three possible reasons that contributed to the encounter of serendipity:

- The office environment is resource-rich, including posters, notifications, different online libraries, etc. Where more information is presented to a participant, s/he will have an increased possibility of experiencing serendipity;
- It provides an interdisciplinary social setting. The work settings for the participants was interdisciplinary, so a participant from mechanical engineering would be sitting in the same office as colleagues from other backgrounds, such as chemical engineering, architecture, or computer science, etc. Foster and Ford (2003) provided several examples of experiencing serendipity among interdisciplinary researchers, and similar cases were also collected in this study. For example, one participant (P 2) from mechanical engineering accidently learnt about a new image-searching engine "TinEye" during a break, when talking with a colleague from computer science, which helped him locate the

resource literature for an image he had used in his writing. Another participant (P 6, design background) also encountered useful information about using Nvivo to help his data analysis, of which he was not previously aware, from a casual conversation with a colleague with an HCI background;

- It makes it easy to get access to different resources. Consider the following example. One participant (P 7) sent us a case in which he happened to encounter a method of "histogram equalization" from a blog when he was browsing the Internet in the office. He then conducted a further search into this method (by referring to Wikipedia and other relevant literature), which enabled him to understand this method and recognise its value (i.e. it could be used in his own research). Compared to other environments, such as a laboratory or a café, it is obvious that the accessibility of resources (e.g. licenses to libraries) impact a participant's judgement of the value of the encountered information.

(3) Status. Status here refers primarily to a participant's commitment to certain on-going activities. Three different types of personal status were identified from the study:

- Leisure: the participant was in a relatively relaxed and open state, such as travelling, playing games, flicking through interesting books or browsing online information, etc. This was a state in which the participant was in his/her own private time fulfilling his/her own interests.
- Seminar: the participant was attending a seminar, a lecture or a presentation where the participant was a student or a listener.
- Working/Studying: the participant was in an intense and focused state carrying out research-related tasks.

According to Table 5-3, among the three different types of personal status, participants tended to experience serendipity more often during their "leisure" time. Studies show that openness and a relaxed setting can facilitate encountering serendipity (McCay-Peet and Toms, 2015; Sun, et al., 2011). Compared to the status of attending a seminar or working/studying, the participants in the leisure status were more open and relaxed.

5.5.2 Social context

Socialisation has always been considered as a significant factor when it comes to discussions on the role of context (Foster, 2004). In all collected cases, aside from the 37 of 62 cases where the participants had a different personal status, the remaining 25 all took place when they were socialising with others (see Table 5-4).

Table 5-4 Different Socialisation Partners

Socialisation Partners	Number of Serendipitous Cases
Classmate	13
Colleague	3
Friend	3
Unfamiliar	3
Superior	1
Any student	1
Group meeting	1

Table 5-4 illustrates that participants experienced serendipity frequently when they were socialising with different people, ranging from the familiar (e.g. classmates, colleagues) to the unfamiliar. However, an obvious trend which can be identified from the table is that the participants tended to experience serendipity more often during periods of socialisation with their peers (classmates, colleagues and friends), while only one case occurred during contact with a superior. This differs from previous

study (Sun et al, 2011), where several collected serendipity cases under a social context came from communicating with superiors. A possible reason for such a phenomenon is "power-distance", which is a widely-understood cultural difference between the West and the East (Hofstede, 1980), while China has been confirmed as one country with a high power-distance, where students are known to keep a larger interaction distance from their professors (Richardson and Smith, 2007). Therefore, compared to communicating with their superiors, the Chinese participants were more likely to communicate with their peers, leading to more serendipitous encounters.

5.5.3 Internal context

Based on the collected data, internal contexts can be divided into the following three aspects: precipitating conditions, sagacity/perceptiveness and emotions.

5.5.3.1 Precipitating conditions

Precipitating conditions refers to the prior conditions for experiencing serendipity. In this study, I have defined such precipitating conditions from a more subjective perspective, which is in relation to an individual's mind-set, covering *visceral needs*, *conscious needs* and *previous experience/ knowledge*.

- *Visceral needs.* As explained in Section 5.4.2, a visceral need refers to a need that does not exist in a participant's remembered experience, and it is not raised until the moment the unexpected information is encountered, as illustrated in the following example:

> *During a casual conversation with a friend of mine, I was really surprised to know that she was in collaboration with one of my classmates in graduate school whom I had not contacted since our graduation. It was an unexpected piece of information to me and I think there would be a high possibility that I could collaborate with him in the future. (Case 44)*

In this case, the participant was initially unconscious of her need to contact an old classmate. It was not until encountering the unexpected information (i.e. her friend was working with him) that she became aware of such a need (to make contact and perhaps collaborate in the future).

- *Conscious needs.* This indicates that a participant had encountered a need/concern at a previous time (e.g. the need to download a useful paper), but for some reason the need/concern failed to be addressed immediately (e.g. unable to access the data resource). Such a need/concern was resolved when the participant encountered the information unexpectedly, as in the following example:

> *I was doing my own experiments recently and gathered loads of experimental data. However, I was not clear how to deal with the errors of the experimental data. During a fluid lab session, I was demonstrating the experiment procedures to the UG (undergraduate) students when I accidently found a handbook near the lab facilities which explains how to read data and deal with errors. I just felt like I had found a big treasure. I started reading immediately; it is easy to understand and is really helpful. (Case 16)*

Clearly, the participant was conscious of his needs/concerns related to dealing with the errors in his experimental data, to keep his research moving forward. However, unexpectedly, this need/concern was addressed as a result of serendipity.

- *Previous experience/knowledge.* Several participants also reported that serendipitous experiences were triggered by their previous experience or knowledge:

> *Today I was cleaning up my summer clothes. Then suddenly I found my old computer which I hadn't used for a long time. When I turned it on, I saw a picture folder which contained all the pictures I took during my UG and this file was the only copy I had! They were really precious memories to me and I had never thought they would come to me in such an unexpected way! (Case 20)*

> *I taught my students an old Chinese poem yesterday in my class. And today I just saw*

the same poem on one of my friends' Wechat shared pages. What a surprise! Even though we were far away from each other, we still shared the same feelings from the old poem! (Case 60)

In Case 20, the accidental discovery of the old pictures helped the participant to recall precious memories of her university life, so the sense of serendipity felt by the participant was mainly based on her previous experience. The situation is similar to the information encountered coincidentally (i.e. recalling the same poem) in Case 60. These cases demonstrate that the prior knowledge, interests and personal experiences can be recalled in unexpected ways, thereby contributing to serendipitous experiences.

5.5.3.2 Sagacity/Perceptiveness

Unexpectedness and precipitating conditions are two necessary components for making connections. However, sagacity or perceptiveness is also required when making such connections. Heinström (2006) and Rubin et al. (2011) have both considered the act of noticing as an ability to "shift the attention from a primary activity to a clue in the environment", and it is "one of the central elements in incidental information acquisition". Sagacity/perceptiveness is also reflected in this study, as in the following example:

It was annoying to search for academic articles in China. I used Go-Agent before, but it no longer works in China. Later on, I encountered serendipity on three different occasions and finally selected the best option for searching academic articles: the first was through the Wechat chatting group, where many members in the group proposed different solutions; the second searching method I learnt was in the laundry room where I met a post graduate student and he talked about some software which he told me that even foreign students use to search for academic articles; and the third method was learnt in the canteen where I met my senior and discussed the issue with him and he proposed a solution to me. (Case 30)

The participant experienced different information on three different occasions, and

during the interview he emphasised how it was the concatenation of the three experiences that made him consider the whole matter as one piece of serendipity. The sagacity/perceptiveness in this case helped the participant to make connections between a conscious need (strategies to search for academic papers) and the unexpected encountering of information on three different occasions. Apart from this example, during the two-week study, the number of serendipitous cases collected from each participant ranged from one to thirteen. Such differences in sagacity/perceptiveness are consistent with the discovery by Erdelez (1997) that serendipitous encounterers can range from super-encounterers to occasional encounterers.

5.5.3.3 Emotions

When a connection is made between unexpectedness and the precipitating conditions, participants switched their attention from the current task to an evaluation of the serendipity they had encountered. Once the value of an encountered experience is acknowledged, serendipity occurs. The term "valuable outcome", which has been highlighted in previous studies (Makri and Blandford, 2012a; McCay-Peet and Toms, 2015; Rubin et al., 2011), is also manifested in this study as follows:

(1) Finding answers to a prior problem or concern. This is most relevant to those serendipities with conscious needs, where the participants are aware of their needs but the answers to prior problems were obtained in an unexpected way.

(2) Providing a potential solution to a need or simply discovering other needs. This is most relevant to visceral needs, as in the example reported in Case 44, where the unexpected information from the participant's classmate may be useful in the future, but whether or not a desirable outcome can be reached in the future

is still unknown.

Such a value is identified as “substantial value”, as it is most relevant to those need-oriented serendipity cases (either for conscious needs or visceral needs), and it is of substantial benefit for helping participants to address their needs or concerns.

However, apart from “substantial value”, I also identified from the study “emotional value” – which highlights the role of emotion– especially in those cases categorised as previous experience/knowledge-oriented. In both Case 20 and Case 60, the unexpected information triggered a huge emotional response from the participants, and it was because such “emotional value” was so compelling that they came to conclude that the encountered experience was an episode of serendipity.

It should be noted that “emotional value” usually accompanies “substantial value” in need-oriented serendipity cases, such that finding solutions to previous conscious needs, or finding potential benefit from visceral needs often accompanies a positive emotion. However, “emotional value” also functions independently in the experience of serendipity, such as in the previous experience/knowledge-oriented cases (e.g. Case 20 and Case 60) in the collected data. Previous studies have pointed out that positive emotions can result from serendipity (McCay-Peet and Toms, 2015; Sun et al., 2011). Nevertheless, few studies have investigated how the role of emotional value is embedded in the process of encountering and perceiving something as serendipity.

5.6 Discussion

5.6.1 A context-based model of serendipity

Based on this empirical study, I identified contextual factors during the processes of serendipity, and further developed a context-based model, as illustrated in Figure 5-

4. This model consists of two parts: 1) In the centre of the model are the three main processes in experiencing serendipity, including encountering unexpectedness, connection-making and value; 2) the impact of external context, social context and internal context on each process during a serendipitous encounter.

The participants' experiencing of serendipity began with encountering unexpectedness, either in an unforeseen way and/or in the unexpected content of the encountered information. Connections are then made between the encountered information and the precipitating conditions (i.e. visceral needs, conscious needs or previous experience/knowledge) of the participant. Once the value of the encountered information (i.e. substantial value and/or emotional value) is recognised by the participants, serendipity occurs.

Each process for encountering serendipity is impacted by context. The main impact of external context and social context is that they are the stimuli for encountering unexpectedness. Such unexpectedness would then lead to connection-making by combing the precipitating conditions, including visceral needs, conscious needs or previous experience/knowledge, which are the internal contextual factors of an individual.

The connection-making process depends mainly on the information encounterer's sagacity/perceptiveness. However, there is also the situation in which connections are provided by external variables, such as from an interaction partner during a period of social contact. One of the participants (P 14) sent us a case concerned with "how to prevent falling asleep while driving" when she was discussing something with her friend, when she unexpectedly received information from her friend that passengers may read books aloud during the journey, which may interest the driver and prevent

boredom. The participant considered the received information to be serendipitous because it was both an unexpected idea and she also applied the idea to her own driving, which turned out to be quite useful. In this case, the participant's process of making connections was simplified thanks to her friend's suggestion. Thus, I argue that the external or social context can sometimes facilitate the process of making connections. Furthermore, the interaction time and activity may also prohibit making connections or evaluating the encountered value. One participant (P 6) reported that he had unexpectedly encountered some information relating to his own research while teaching undergraduates in the laboratory. However, he was too busy answering the problems from the students to think further about the encountered information. It was not until he received the daily reminder that he was able to recall the encountered information from earlier in the day, at which point he became aware of the value of the same. In this case, I argue that the participant's status of being busy (e.g. answering questions from students) may have prohibited his immediate recognition of, or attention to, serendipity at the time at which it occurred.

Emotion played an important role in recognising the value of any encounters. Makri and Blandord's (2012a) study found that a good mood may help participants to exploit the value of an encounter, while a bad mood may impede such exploitation (p. 694, case UD1 and AD4). In this study, I identified the value of an encounter as a substantial value and an emotional value, and particularly in any previous experience/knowledge-oriented cases (e.g. recalling good memories, as in Case 20), the emotional value would directly lead participants to consider the encounter as an episode of serendipity. Another point which should be noted is the relationship between emotions and sagacity/perceptiveness (the broken arrow line in Figure 5-4).

Research from the fields of psychology and neuro-science has revealed that emotions can also impact an individual's cognitive processes, such as their memory, decision-making, attention or learning (Schupp et al., 2006; Lerner et al., 2015), which may thus further influence how connections are made between any unexpected information and the precipitating conditions.

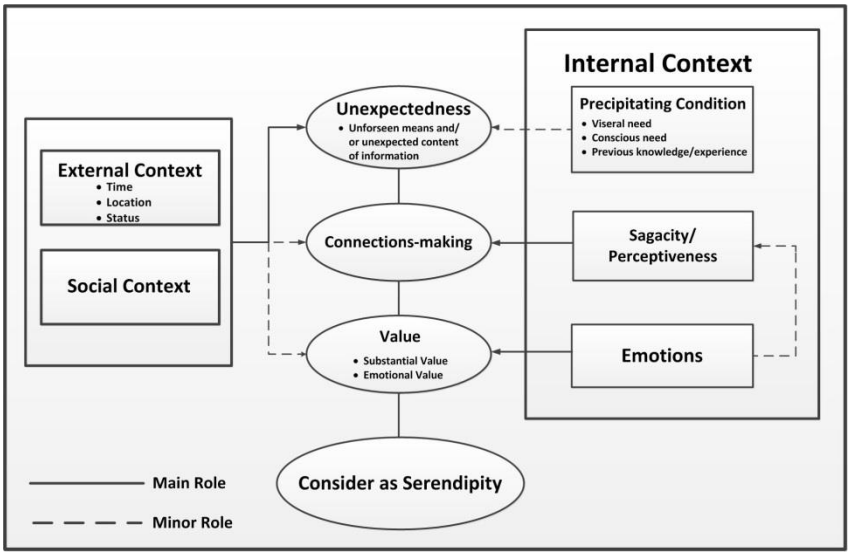


Figure 5-4 A context-based model of serendipity

5.6.2 A comparison with previous model

There are two main differences between this updated model and the previous model (Sun et al., 2011). The first is that the process of connection-making is actually pervasive in all cases of serendipity. In the previous model (Figure 5-1), two different levels of abstraction are identified that can lead to serendipity, and connection-making is the second level of abstraction that can sometimes result from the first level. However, by identifying the internal context of the precipitating conditions in this new empirical study, I have found that the original first level of abstraction also involves a connection-making process. For example, a previous case was idenfitied as the first level of abstraction where the participant accidently noticed some interesting

module codes when he was looking at the whiteboard during a workshop. The participant was looking for module information from different sources (leaflets, the Internet) at that time, and this new module turned out to be useful. The authors previously considered there to be no connection-making during this case and identified it as non-activity-based, unexpectedly valuable information from unexpected sources (a category of the first level of abstraction). However, according to this new identification of internal context, this participant had a conscious need in mind (always looking for interesting module information), and the unforeseen means by which he obtained the new module source from the whiteboard during the workshop (unexpectedness) led to a connection between this unexpectedness and the conscious need, and when he finally recognised the value of the encounter (a useful module), serendipity occurred. I believe this new identification of internal context can better help researchers to understand the connection-making process during serendipity.

The second difference between this updated model and the previous model is that it demonstrates the impact of different contexts during each process of serendipity. The previous model mainly discussed the role of context in encountering serendipity from three aspects: the role of people (active or less active), the role of temporal factors, and the role of the environment (i.e. the working environment, places, and the changing environment). It is directly adopted from Schmidt's context model (2000) to denote the relationship between the role of the individual and their context in serendipity, but failed to explain how these contextual factors affected the different processes of a serendipitous episode. This element is complemented in this empirical study through a new identification of contextual factors:

1) The role of people was further identified by explaining the internal context of the precipitating conditions (i.e. visceral needs, conscious needs and previous experience/knowledge), sagacity/perceptiveness, and emotions. The precipitating conditions form a premise for each individual to make connections when unexpected encountering happens, sagacity/perceptiveness impacts mainly on the process of making connections, while emotion can affect a participant's recognition of the encountered value, and may also impact an individual's sagacity/perceptiveness when it comes to making connections.

2) The role of temporal factors and the role of the environment were further identified by defining the external context (i.e. time, location, and personal status) and social context. The external context and social context are the stimuli for unexpected encountering, and they also affect a participant's connection-making (e.g. facilitation) and recognition of the encountered value (e.g. being busy).

The proposed new context model verified serendipity as "the product of context" (Foster and Ellis, 2014, p. 18), and these external, social and internal contexts play different roles and are interwoven throughout the encountering process of serendipity.

5.6.3 Implications of the updated model

5.6.3.1 An extension of the existing framework

Three processes for experiencing serendipity were identified from the study: "unexpectedness", "connection-making" and "value". This identification is similar to the framework proposed by Makri and Blandford (2012b), in which they considered "unexpectedness", "insight" and "value" as the three key elements with which to evaluate serendipity, and where the "making of the connection itself involves an

amount of insight" (p. 714). The new contribution to this framework is that, based on this empirical study, the three processes are further expanded. I have identified three different channels that lead to "unexpectedness", the different situations of connection-making between the encountered unexpectedness and the precipitating conditions of visceral needs, conscious needs and previous knowledge/experience, and I have also identified value as substantial value and emotional value.

The identification of "unexpectedness" in this work is similar to the work by Foster and Ford (2003), in which they classified four different categories of serendipity. My identification of "unforeseen means of encountering information" is similar to Foster and Ford's third category, "the unexpected finding of information the *existence* and and/or *location* of which was unexpected, rather than the value" (p.332), and the identification of the "unexpected content of the encountered information" is similar to Foster and Ford's fourth category, "the unexpected finding of information that also proved to be of unexpected value: (a) by looking in "likely" sources"; (b) by chance" (p.332). However, this work also goes beyond their framework by identifying the processes of "connection-making" and "value", which is not discussed in their paper. By expanding the different processes of serendipity, I have found it is possible to give a new classification of the different categories of serendipity. Table 5-5 categorises the cases listed in the previous sections of this chapter and:

- For a conscious need/concern, it is more the unforeseen means of encountering the information that results in the participants' sense of unexpectedness, and brings both substantial value (e.g. an answer to the concern) and emotional value (the positive emotion when the need is addressed), as identified in Case 7, Case 44 and Case 30.

- For a visceral need/concern, it is usually the unexpected content of the encountered information that leads to participants' feelings of unexpectedness, while it also brings both substantial value (e.g. finding a possible solution for a visceral need) and emotional value (a positive emotion), as identified in Case 6 and Case 44. However, sometimes the unforeseen means of encountering information may also play a role in leading to unexpectedness, as identified in Case 1.
- For previous experience/knowledge, it often results in emotional value, and both the unforeseen means of encountering the information and the unexpected content of the encountered information have the potential to bring a feeling of unexpectedness to participants, as identified in Case 20 and Case 60.

Table 5-5 Classification of serendipity cases based on the framework

Serendipity Cases	***Connection Making***					***Value***	
	Unexpectedness		***Precipitation Conditions***				
	Unforeseen Means of encountering Information	*Unexpected content of encountered information*	*Visceral need*	*Conscious need*	*Previous Experience/Knowledge*	*Emotional Value*	*Substantial Value*
Case 7	✓			✓		✓	✓
Case 6		✓	✓			✓	✓
Case 1	✓	✓	✓			✓	✓
Case 16	✓			✓		✓	✓
Case 44		✓	✓			✓	✓
Case 20	✓				✓	✓	
Case 60		✓			✓	✓	
Case 30	✓	✓		✓		✓	✓

5.6.3.2 Design strategies based on the identified contextual factors

Björneborn (2017) argues that:

> We cannot design environments always leading to serendipity – as serendipity is a highly subjective and situational phenomenon. But affordances for serendipity can be engineered……Serendipity may thus be intended by designers, but must always be unplanned by users (p. 1068).

From this empirical study, the identified contextual factors of external context, social context and internal context helped me to look into the role played by context during the different processes of serendipity, and thus provide possible implications for designers to design affordances that can "engineer serendipity". These include: (1) consider participants' status. Participants during the study tended to experience serendipity more often in the afternoon than in the morning, especially when they are in their leisure time. When organising activities that aim to facilitate serendipity (e.g. free discussion seminars, using information systems to make recommendations), it is better to take participants' status into consideration and arrange such activities during their leisure time; (2) consider locations. This study has found that places with rich resources, such as a physical resource (e.g. licensed online libraries) or a socialisation resource (e.g. interdisciplinary offices), and with easy accessibility provide greater potential for participants to encounter serendipity; (3) try to create social networks. Social context is a significant stimulus, where participants can encounter unexpectedness which may result in serendipity, so attention should be given to such context to help to "engineer serendipity"; (4) design environments that are both diverse and conspicuous. Motivated by an understanding of the internal context of pre-conditions and sagacity/noticeability, I also suggest the design of more diverse

and noticeable environments to encourage the occurrence of serendipity. Current information technologies, such as recommendations, personalisation and visualisation may consider this aspect to facilitate the occurrence of serendipity; (5) combine emotional design. As identified in the study, emotion plays an important role during the process of encountering serendipity, so an element of curious and/or interesting information may help to encourage the occurrence of serendipity. Those designs for joy, surprise and/or other emotional design strategies can also be applied to serendipitous design strategies.

5.7 Conclusion

This chapter presents an empirical study where the social media platform "Wechat" was employed as a research tool to investigate the phenomenon and occurrence of serendipity in the context of information research among a group of Chinese scholars.

Based on the collected data, it can be found that current understandings of serendipity, which have been constructed mainly on the basis of Westerners, also applied to the Chinese participants. They also experienced serendipity according to the three main processes of encountering unexpectedness, connection-making and recognising the value. The three processes are further expanded in this empirical study. Unexpectedness is encountered by any unforeseen means and/or the unexpected content of information encountering, where connections are made between the unexpectedness and the precipitating conditions of visceral needs, conscious needs, or previous experience/knowledge. Ultimately, either a substantial

value or an emotional value prompts the individuals' recognition of serendipity. This expanded framework can also help to classify the categories of serendipity.

The role of context in experiencing serendipity has been further investigated. Compared to the previous context model, the updated context-based serendipity model better demonstrates the different interactions and influences of the external context, social context and internal context during the different processes of serendipity. In particular, this study found that the role of emotions should not be considered only as an outcome of serendipity, but it should also be embedded in the process of encountering serendipity, which is an issue that has been largely neglected in existing studies.

CHAPTER 6

USER STUDY3 —— OBSERVE SERENDIPITY IN CONTROLLED RESEARCH SETTING: THE ROLE OF EMOTIONS IN SERENDIPITOUS ENCOUNTERING

Research questions addressed in this chapter:

What is the role of emotion during a serendipitous encountering?

[1] How can a participant's emotions be evoked and captured in a research setting?

[2] How should a research environment be designed to facilitate participant's encountering of serendipity?

6.1 Introduction

In Chapter 5, a new context-based model of serendipity is constructed where emotion is considered as an element of internal context that played a role in the process for a participant to encounter serendipity. This chapter further explores how emotions can impact the process of a serendipitous encounter.

Following a review of existing studies on serendipity, I have found that few have paid attention to gaining an understanding of the role of emotions in the process of serendipity. Emotions are often regarded as a positive outcome resulting from

serendipity, rather than as a potential factor that may affect the occurrence of a serendipitous encounter (Sun et al., 2011). In their study, Rubin et al. (2011) discovered that "surprise" played a role in accidental encounters, but they failed to make any further exploration into how any such "surprises" would have an impact on the experience of serendipity.

This chapter presents the results of an empirical study undertaken with participants in a controlled laboratory, wearing physiological sensors and monitored by an Eye-tracker, with the objective of investigating the role of emotions during serendipitous encounter. The chapter is constructed as follows: Section 6.2 introduces the research background for the impact of emotions on experiencing serendipity; Section 6.3 proposes the research questions and introduces the developed application as a research platform; Section 6.4 illustrates the methodology applied in the study. The research findings and discussion are presented in Sections 6.5 and Section 6.6 respectively, while Section 6.7 is the conclusion.

6.2 Background

6.2.1 The study of emotions in information science

The role played by emotions in information science has been recognised by an increasing number of researchers in recent years. For example, González‐Ibáñez, Shah, & Córdova‐Rubio (2011) pointed out that two questions are confronted in the research area of information science: "[Q1] do [people] feel a certain way because they find (or do not find) information? [Q2] Or do [people] find (or do not find) information because they feel a certain way?" While current research into emotions in

information science has focused primarily on the first question, which considers how individuals' affective states or emotions are dependent on the information seeking or retrieving processes, the second question has been largely ignored. This highlights how is also possible that emotions may influence individuals' information seeking behaviours. In a similar vein, Lopatovska and Arapakis (2011) reviewed the causes, effects and correlates of emotions in different information contexts, including computer games, online searches, digital libraries, etc. They argued that the study of emotions in the domain of HCI is still in its infancy, and suggested that it is worth broadening and enriching the research on the affective factors inter-disciplinarily. In the third version of his book, Case (2012) reviewed nine general applied information seeking models, and among them, only Kuhlthau's information searching process (ISP) model (1991) highlights the constructive role of feelings and moods in an individual's finding and evaluating information. However, in the latest and fourth version (Case and Given, 2016), they have updated the collected models to twelve, and apart from Kuhithau's model, another two models (i.e. the Foster model (2004) and the Robson and Robinson model (2013)) also take much account of the role of psychological factors (e.g. feelings, emotions) on an individual's information behaviour. Furthermore, in the conclusion part, Case updated two additional implications when compared to his third version, one of which is that "people's experience of information is not independent of their emotions", and he also argues that "emotional connections must be taken into account in analysing individuals' information behaviours."

In the past decade, a number of empirical studies have also demonstrated the influence of emotions on information behaviours. For example, Fulton (2009) found

positive emotions or feelings were a primary motivation for beginning the information seeking processe. González-Ibáñez & Shah's (2012) study showed that both positive and negative emotions can potentially affect an individual or a team's information seeking process. In a specifically designed task, participants who received positive emotional stimuli performed better than those who received negative stimuli. Porat & Tractinsky (2012) found that the aesthetics and usability design of a web store may impact a visitor's emotions (e.g. pleasure, arousal), which in turn may influence the visitor's attitude toward the store. Savolainen (2014) conducted critical conceptual analysis and found that emotions and feelings could have an impact on different information seeking categories, namely, starting, expanding, limiting, terminating and avoiding. More specifically, positive emotions were often found to be associated with users starting and expanding their information seeking, while negative emotions had motivational power in all five of the information seeking categories. Wu's (2015) study found that children searching for information under a positive emotional status (entertainment) could better endure information uncertainty compared to those searching for information merely with the aim of solving a problem. Wang, Zhou & Jin's (2017) study showed that the positive emotion produced by different cues (i.e. information, ambience and socialisation) is an important drive for an individual's sharing of information on microblogs. These findings provide sufficient proof that the role of emotions deserves to be acknowledged in the area of information science.

6.2.2 Study of emotion in serendipity

Serendipity is considered by other researchers as "chance encounters" (Toms and McCay-Peet, 2009), or "incidental information acquisition" (Heinström, 2006). With

regard to serendipity research, emotions are one of the most influential factors resulting from serendipitous encountering. For example, Sun et al. (2011) found that a significant impact of serendipity was the emotionally positive effect on their participants, while Yadamsuren and Heinström (2011) found that there would be different emotional reactions when people were confronted with incidental access to online news. In addition, Rubin et al. (2011) discovered that positive emotionality (e.g. happy endings, memories of past experiences) is part of the perceived value brought by the fortuitous outcome of a serendipitous experience, and they further found that "surprise" is somewhat related to serendipity, although its role remains to be identified in future studies. Similarly, the collected cases from Makri and Blandford (2012a) found that there are some emotional links between serendipity and a happy feelings (EDR 1), while McCay-Peet and Toms (2015) also found how emotions can help to highlight "triggers" and facilitate serendipity (P3).

Nevertheless, all the aforementioned studies failed to have a systematic investigation into the role of emotions in experiencing serendipity. To the best of my knowledge, the only reported study which examined the psychological factors relating to serendipitous encountering was carried out by Heinström (2006), who also considered serendipity to be "incidental information acquisition". Based on a survey of 574 students in grades 6 to 12, from a broad range of curriculum topics, her study found that those students with feelings of confidence, certainty, and satisfaction tended to be more likely to experience serendipity. However, as argued in the paper, the survey was based on participants' self-reporting questionnaires, so their responses to the questions were grounded in their own memories, rather than in a realistic occurrence of serendipitous environments. The particular features (e.g. physical signs,

follow ups) found during the process of serendipity failed to receive further investigation.

6.3 Research Design and Preparation

6.3.1 Two methodological challenges

The main objective of this chapter is to examine the role of emotions during serendipitous episodes under the conditions of a controlled laboratory setting. To achieve the objective, the design of the study must begin by considering two methodological challenges:

Challenge 1: to evoke and capture a participant's emotions in a research setting

There are a number of research methods available which can stimulate participants' emotional responses (e.g. videos, audio, images and games). This study applied a game situation where participants were invited to a self-developed sketch game application, primarily because: 1) games have long been used in psychology and HCI studies in light of their natural ability to evoke emotional reactions (van Reekum et al., 2004), and the employment of games is conducive to researchers to judge the value of emotions, where a successful game result is usually associated with participants' positive emotions, while a failed game result can lead to negative emotions; 2) compared to other methods (e.g. asking participants to finish a given search task), the employment of a game is a more leisurely and fun way to help participants to become relaxed, as it has been proven that serendipity tends to be more easily encountered when people are in a relatively relaxed state (Sun et al., 2011; McCay-Peet and Toms, 2015).

Two components that are widely used to measure the aspects of emotion are emotional valence (i.e. positive and negative) and arousal (Tkalčič et al. 2016). While the emotion valence can be obtained through design of the game, emotional arousal in this study is captured by electrodermal activity (EDA). Human bodies can act as conductors for electricity and sweat is produced through glands in human hands, which in response to emotional and cognitive stimuli. Once an individual experiences stimuli, the salty sweat produces in the gland, which increases body conductivity and lead to a higher flow of electricity (Roth, Dawson, & Filion, 2012; Lazer, 2016). EDA is such a measurement of the flow of electricity through the skin, and the EDA sensor applied in this study uses a pair of electrodes on the fingers to measure the conductivity between two points (Figure 6-1).

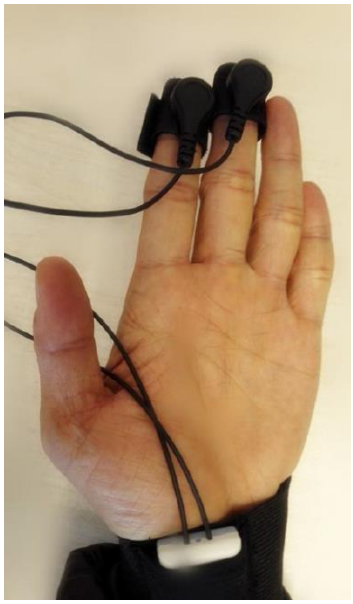


Figure 6-1 EDA sensor applied in the study

Skin Conductance Level (SCL) and Skin Conductance Responses (SCRs) are two indicators of autonomic arousal collected by EDA. SCL shows the overall "tone" of the background skin conductance level, and it is usually used to measure participants'

basal skin conductance over a period of time. SCRs show the "phases" or abrupt increases in skin conductance, and it is usually used to represent the transient and rapid fluctuations of skin conductance levels caused by stimuli (O'haire et al., 2015). Both indicators can be used to assess participant's physiological arousal, and help researchers to link such arousal with participant's emotions like frustration, fear, sadness, etc. (Scheirer, Fernandez, Klein & Picard, 2002; Cacioppo et al., 2000; van Reekum et al., 2004; Ravaja, Turpeinen, Saari, Putonen, & Keltikangas-Järvinen, 2008).

In addition, an Eye-tracker is employed in this study to help record participants' interaction processes with the game. The combination of Eye-track will help researchers to identify the scenes where participant experiences emotional stimuli based on the collected data by EDA, and thus to better understand their emotional responses at that specific time.

This part will be further introduced in Section 6.5.1.

Challenge 2: to design a research environment to facilitate participant's encountering of serendipity

It is also a challenge to define/design research settings in which there is potential for participants to experience serendipity. This research applied a "Wizard of Oz" approach in a self-developed game application according to a collaborative-filtering based algorithm, which was developed with the aim of supporting the potential occurrence of serendipity in recommendation settings.

The collaborative-filtering based algorithm is self-developed based on the understanding of serendipity in information research. The core of the algorithm is to provide users with information with a small possibility of being discovered, but with a

significant relevance to the user. The term "small possibility" applies to situations when information is relatively difficult to be discovered by participants and where such information would probably bring a sense of "unexpectedness" to the user, while the "significant relevance" is based on a user's profile or search history, which could probably help to lead users to a "valuable outcome" (Ge, Delgado-Battenfeld, & Jannach, 2010). A detailed description of the algorithm will be introduced in next chapter.

6.3.2 A sketch game as the research platform

A self-developed game was employed in the presented study as a platform to provide potential serendipitous encounters for participants. It is a sketching game which was developed based on Android 5.0, and the system is implemented with a Client-Server structure, where users can interact with the application on the Client side while the Server side "fetches" requests from the client end and returns corresponding results to users. A colour recommendation system, an image retrieval system and a game system were constructed as three subsystems for the server, in addition to a self-structured image database, which contains over 2,000 pictures.

The developed game involves the following stages (Figure 6-2):

- Memorising and sketching

Each participant was given a picture in the very beginning for observation. Participant was then asked to layout the colour features of the picture based on the memory. A time clock is set during this stage where the maximum observation time is 30 seconds and the maximum sketch time for each participant is 120 seconds.

- Retrieving

When a participant finishes sketching, a group of 30 images is displayed to the participant for retrieving whether or not his/her drawing picture was contained in the provided pictures. If the picture is contained in the group, he/she may click on the picture to pick it out. Or the participant only needs to click “Next” button.

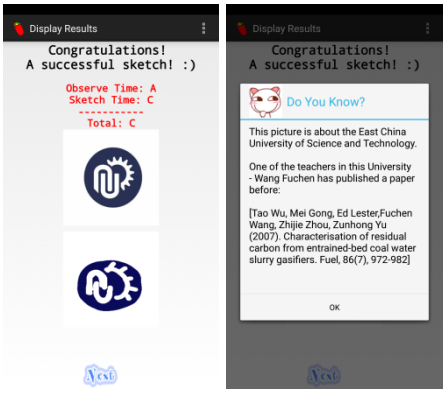
- Sketching result

Participant’s finial sketching result is provided after retrieving. A winning game means the participants has successfully retrieved the drawing picture, and then he/she will be given a game score based on the the observation time and sketching time. Otherwise, the participant will be noticed he/she has failed in the sketching.

- Providing picture information

The last stage of the application provides participants the related information of the picture, in regardless of whether or not the participant has made a successful sketching.

(a) (b) (c)



(d) (e)

Figure 6-2 Different stages of the designed sketch application: (a) Memorised picture; (b) Participant's sketching; (c) Retrieving; (d) Sketching result and game score; (e) provided picture information

6.3.3 Applying Wizard of Oz to the game

A "Wizard of Oz" approach was applied to the study according to the proposed serendipitous algorithm (will be introduced in detail in the next chapter). The concept of the "Wizard of Oz" approach is not new in serendipity studies, as both Pontis et al. (2015) and Kefalidou and Sharples (2016) have used this method to investigate their developed serendipitous systems.

The potential serendipitous encountering which exists in this game application is based on the relationship between the picture provided at the beginning of the sketch and the text information given after the sketch has been completed. By applying the "Wizard of Oz" approach, the researcher acted as a constructor to bridge between the provided picture and the text information according to the participants' academic information, and thus three different experiment conditions were set for the relationship between the picture and the text information.

Condition 1: socialisation-based information

This relationship between the picture and the provided information was designed on the basis of the socialisation relationship between a participant and his/her supervisor's co-author. In more detail, we randomly selected a co-author from the participants' supervisors' online publications, initially providing each participant with a picture of the logo of the institution in which the participants' supervisors' co-author works. The socialisation relationships were not provided to participants until they had finished their sketches. Such design was inspired by McCay-Peet and Toms's (2015) suggestion that 'serendipity may be facilitated by creating or immersing oneself in environments that are trigger-rich or contain the perceptual cues that act as a catalyst for serendipity', and 'the prepared mind primes individuals to recognize potential triggers, helps them make connections, and follow up on them'. The socialisation-based information, which is on the basis of a participant's own experience with his/her supervisor, is in accordance with the concept of a "prepared mind" and would act as a "trigger" to facilitate serendipity. Taking participant 1 as an example, the provided information was as follows:

> *"This picture is about the East China University of Science and Technology. One of the teachers in this university – Wang Fuchen [sic] has published a paper: Tao Wu, Mei Gong, Ed Lester, Fuchen Wang, Zhijie Zhou, Zunhong Yu (2007). Characterisation of residual carbon from entrained-bed cold water slurry gasifiers. Fuel, 86(7), 972-982." (Fig.1, picture e)*

The information both introduces the institution, and gives details of a published paper which contains the names of both the supervisor of the participant (Tao Wu) and the co-author (Fuchen Wang). I hoped that the supervisor's name in this case would act as a "perceptual cue" or a "trigger" and would lead to the occurrence of serendipity.

Condition 2: research background-based information

This relationship between the picture and the provided information was designed on the basis of a participant's research background. For example, a picture of "London's Tower Bridge" was provided to participant 1, whose research is concentrated on Chemical Engineering, as well as the following information:

> *"This picture is about London, in the United Kingdom, where the 18th International Conference on Chemical, Ecological and Environmental Sciences (ICCEES 2016) will be held next January."*

The information provided facts about an upcoming conference in the participant's research field. Similarly, the given conference information would act as a catalyst for experiencing serendipity.

Condition 3 (controlled condition): conventional information

A controlled condition is selected as a comparison of the previous two conditions. It was more conventional, without employing any "triggers" or "prepared mind" techniques between the provided picture and the text information. A cover picture was chosen from a research website (www.nature.com) and the provided information was followed by an introduction to the picture on the website:

> *"This picture is the cover image of 'Nature' from July 16th, 2015. The relevant information is: In collaboration with Scientific American, Nature takes a look at modern movements in teaching science, technology, engineering and mathematics (STEM). By applying the principles of twenty-first-century learning, educators should be able to produce scientists better prepared for the modern, multidisciplinary workforce and a more science-literate populace in general."*

Take condition 1 for example, the algorithm is embedded into the sketch game based on three assumptions:

- Assumption 1: For each PhD student, their supervisor's information is a large weight attribute in their personal profile.
- Assumption 2: For each PhD's supervisor, the co-author from their publications is a large weight attribute in the supervisor's profile.
- Assumption 3: For each co-author's personal profile, their working institution is a large weight attribute.

During the game session, each PhD student supposed to be provided with the information of their supervisor's co-author's institution. Figure 6-3 shows the design of the study including how the proposed algorithm is embedded into the game-based application and the sketch game process (Condition 1 and Condition 2). The game started by providing a picture which showed the institution badge (Figure 6-2-a), and the serendipitous information of the picture was provided to the participant when the drawing is finished (Figure 6-2-e). The given information related to the picture includes two levels: (1) the introduction of the institution; (2) the publications of both the participant's supervisor and the co-author, as is shown in Figure 6-5-a.

As a comparison, each participant was also given the pictures that without the serendipitous information from the proposed algorithm (Condition 3, see Figure 6-4). Two cover pictures from the "Nature" website (www.nature.com) were selected to the participant, together with the description of the picture on the website, see Figure 6-5-b as the example. Such provision is a conventional way to introduce the relationship between the picture and the information (pic-and-info).

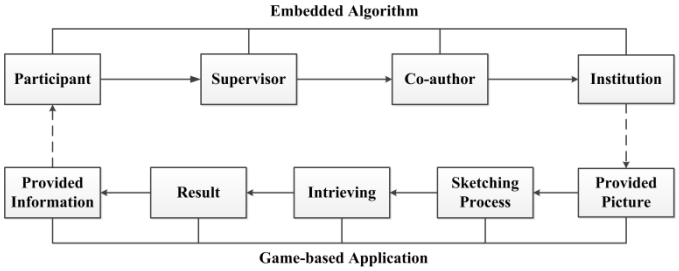


Figure 6-3 Process of the study and the embedded proposed algorithm

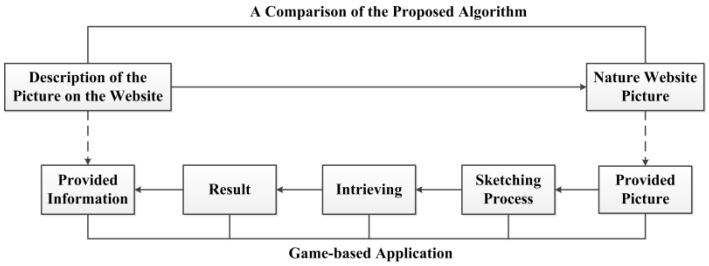
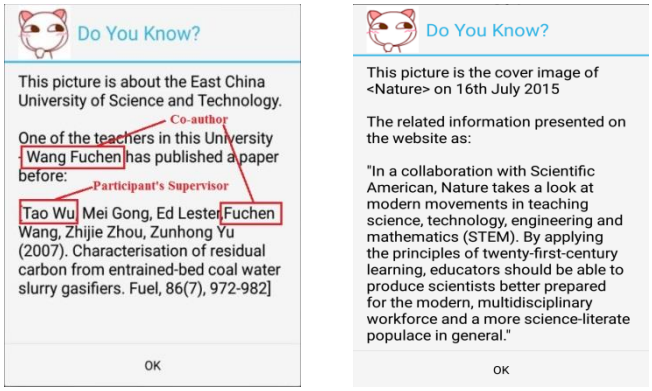


Figure 6-4 A comparison of the proposed algorithm



(a) (b)

Figure 6-5 Provided information: (a) designed algorithm; (b) information from the nature website

6.4 Methodology

6.4.1 Participants

The study was first approved by the ethics committee of the university, following which 26 PhD students from the university (13 males and 13 females) were invited to participate in the study, with a mean age of 28.1 (SD=2.6). All participants' academic

information can be found on the university's website, including their research background and supervisors' information.

6.4.2. Procedure

(1) Pre-introduction. Each participant was first introduced to the sketch application. It took approximately ten minutes for each participant to become familiar with the game, and the study did not begin until they had finished drawing at least two pictures.

(2) Preparation for the EDA skin conductance sensor. When each participant had become familiar with the drawing-game application, he/she was asked to sit in front of a fixed mobile phone and put on the EDA skin conductance sensor. After the participant finished calibrating the Eye-tracker and had adjusted him/herself to a comfortable position, with even breathing and balanced emotions after a period of rest (i.e. when the SC signal trends were constant), the study began with a 3-minute basic skin conductance test to collect the baseline characteristic values of the participants.

(3) Drawing, and filling in the questionnaire. Participants were then asked to start their drawings and each of them was given six pictures to sketch (with two pictures for each condition). When a participant had finished one drawing, he/she was asked to fill in the questionnaire in relation to the given text, based on the four dimensions of being "unexpected", "interesting", "relevant" and "beneficial". The design of the questionnaire was initially based on the study by Makri and Blandford (2012b), where the dimensions of being "unexpected", "interesting" and "beneficial" were chosen as measurements of serendipity. Pontis et al. (2015) employed these measurements and further highlighted in their latest study that "relevance" should also be an important element to consider when measuring serendipity. Consequently, "relevant" was also

added as a dimension in the questionnaire. Each dimension was measured/scored on a five-point Likert scale, rated as 1 for 'strongly disagree' and 5 for 'strongly agree'. After completing each questionnaire, participants were given 30 seconds to recover with his/her eyes closed until commencing the next drawing.

(4) Post-interview and thinking-aloud. When participants had finished each of the six drawings, and the EDA skin conductance sensors had been removed, they were invited to a post-interview. The interview aimed to investigate the participants' reasons for their responses to the questionnaire, their opinions/experiences of the developed application and the research method employed. In particular, a thinking-aloud session was conducted at the end of the interview, where all the collected data (including the EDA data and recordings from the Eye-tracker) was presented to participants to help them recall their sketching processes. Participants were then asked to express their feelings on some specific peaks (potential SCRs) in their EDA data.

6.4.3. Collected data and analysis

Both objective data and subjective data were collected from the empirical study. The objective data includes the collected data from the EDA skin conductance sensors and the recordings from the Eye-tracker during the participants' interaction with the game. More specifically, the EDA skin conductance sensors collected psychophysiological data from the participants, including Skin Conductance Level (SCL) and Skin Conductance Responses (SCRs). The Eye-tracker recorded the participants' interactions during different game procedures (as shown in Figure 6-2), and thus helped to analyse the psychophysiological data accordingly.

The subjective data includes data from the questionnaires and the interviews. SPSS 21 was used to make a statistical analysis of the collected data from the EDA

skin conductance sensor and the questionnaires. The interview data were organised and analysed through the use of an affinity diagram.

6.5 Findings

6.5.1 Emotion responses during the study

6.5.1.1. Differences in SCL before and after drawing

SCL is an important parameter with which to measure emotional arousal in psychology studies (El-Sheikh, 2007). However, individual differences exist in skin conductance signals (El-Sheikh, 2007), so one solution to address this problem is to use the skin conductance rate as an index to represent the change in a participant's SCL, taken from readings made before and after the application of any stimulus (Zhang, Kang, & Jin, 2015). It can be calculated by the proportion of the SCL value with the formula:

$$R = (X_{SC} - X_{base}) / X_{base} \quad (1)$$

where R represents the skin conductance rate, X_{SC} represents the SCL value after the stimulus and X_{base} represents the basic SCL value of each participant before the stimulus. In this experimental setting, if the calculated R is positive, it represents that the drawing has successfully brought emotional arousal to participants; in contrast, if the calculated R is negative, it represents that the designed study failed to bring participant emotional arousal.

Figure 6-6 gives the drop-box of the SCL rate change for each participant's successful drawing and failed drawing. It can be seen that the game effectively changed the participants' SCL values and led to a rising rate for both drawing results.

Table 6-1 lists the detailed SCL rate change under the two conditions of successful drawings (N=92) and failed drawings (N= 64). The mean SCL rate in both conditions shows an improvement, with the mean rate at 2.55 (SD=2.18) when the drawing was successful and 2.24 (SD=2.01) when the drawing was unsuccessful, but no significant differences exist between the two conditions (i.e. successful drawing vs. failed drawing). The change of the SCL rate implies that the game has successfully stimulated the participants' emotions, either with a successful drawing or with a failed drawing solution. Based on the results, two assumptions were proposed relating to the valence of emotions:

- For all the participants in the study, a successful drawing is related to a positive emotion;
- For all the participants in the study, a failed drawing is related to a negative emotion.

I believe such assumptions are appropriate for this study because:

1) According to Tkalčič et al. (2016), in the area of HCI emotions are referred as "affect states that are triggered, last for a short period of time and have several dimensions (e.g. valence and arousal)" (p.5). In this experimental setting, either a successful drawing or a failed drawing will not last for a long time, as each participant needs to finish six pictures and their emotions are triggered by the result of each sketching and will often changes with the different results (i.e. success and failure); therefore it is possible to use the valence of emotion to describe participant's affect state during a short time period after they finish each drawing;

2) Fredrickson (1998) proposed a model to describe the form and function of positive emotions, which includes *joy*, *interest*, *contentment* and *love*. These positive

emotions can help to "broaden an individual's momentary thought-action repertoire, which in turn has the effect of building that individual's physical, intellectual, and social resources" (p.1). Fredrickson (1998) argued that the positive emotion of *joy* can arise by "events construed as accomplishments or progress towards one's goals" (p. 6). Looking back into this designed game study, participants are required to finish their task in a game-setting with a clear goal —— to sketch out the observed picture; therefore in this experimental setting I define the state of "joy" for a participant to successfully finish the task, achieve the goal with accomplishment as a positive emotional state.

3) Negative emotions typically include fear, anxiety, loneliness, guilt, shame, embarrassment, regret, disappointment, sadness, envy, jealousy, disgust, scorn, anger, frustration, and irritation (Parrot, 2014). During the interview session, I did receive participants' feedbacks about these negative feelings when they failed in their sketching. For example, one participant (P 3) argued that if he failed in the drawing, he would be quite frustrated and kept on thinking which part of the drawing led to his failure; another participant (P 11) argued that she would be quite embarrassed if she failed in drawing out a picture which she considered as quite simple. I thus define such a state for a participant failed in achieving a successful result with other negative feelings (e.g. frustration, embarrass) as a negative emotional state.

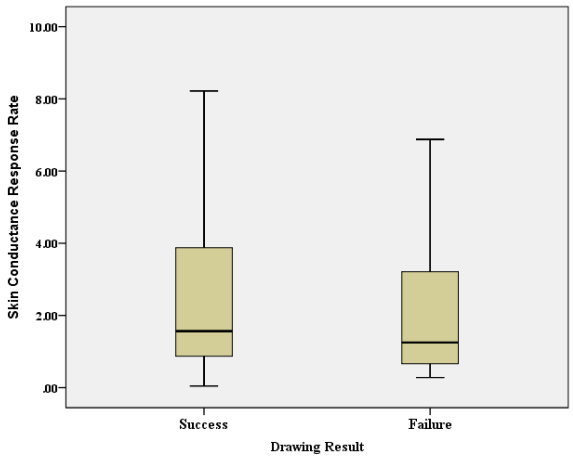


Figure 6-6 Drop-box of participants' SCL rate change for different drawing results

Table 6-1 Participants' SCL rate with the drawing results as a variable

Condition	Number	Mean	SD
Successful Drawing	92	2.55	2.18
Failed Drawing	64	2.24	2.01

6.5.1.2. Discovered SCRs

The potential occurrence of serendipity designed in this study is based on the provided text information under the "Wizard of Oz" approach. Apart from the SCL rate change during the game session, event-related skin conductance responses (SCRs) were also used to analyse the participants' emotional arousal during the text reading process. According to O'haire et al. (2015), if there is SCRs identified, it represents that a participant is experiencing an emotional stimuli at that specific time node. As a result, there were 156 pictures in total for all participants to sketch, but not all the information relating to these pictures was read by the participants. Those pictures which were not viewed by participants were eliminated under the help of collected data from the Eye-tracker. As a result, 83% (N=129) effective reading

pictures were picked out. Table 6-2 demonstrates the discovered SCRs among the effective viewed pictures, with the two conditions of successful drawing and failed drawing. In total, 67% (51 out of 76) of the viewed pictures after a successful drawing (positive emotion) were discovered with SCRs, while 64% (34 out of 53) of the viewed pictures after a failed drawing (negative emotion) were discovered with SCRs. Figure 6-7 lists an example of the SCRs discovered from the study, taken from a participant's (P16) reading of the text after completing the drawing.

Table 6-2 Discovered SCRs under different conditions

Drawing Result	Condition	Number
Success	With SCRs	51
	No SCRs	25
Failure	With SCRs	34
	No SCRs	19

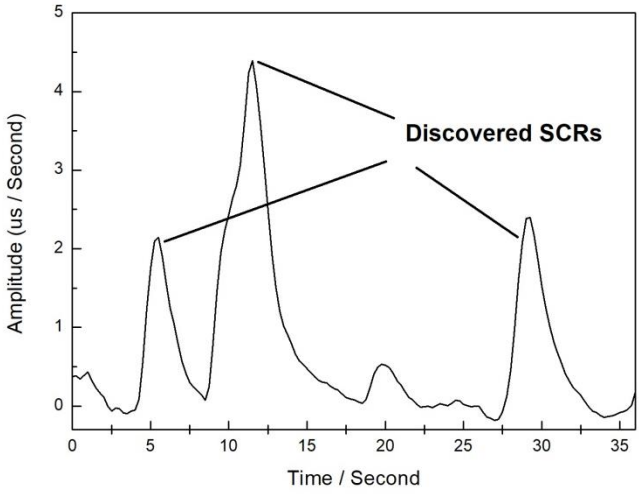


Figure 6-7 Discovered SCRs during reading of the text

6.5.1.3. Participants' experience of serendipity

The evaluation of serendipity was mainly dependent on participants' scoring on the questionnaire. Following the instructions of Pontis et al. (2015), who employed the framework of serendipity evaluation which was developed by Makri and Blandford (2012b), only those scores over 4 on all dimensions of the questionnaire (i.e. unexpected, interesting, relevant and beneficial) were considered as effective cases of serendipity. In particular, when measuring dimensions on serendipity, the information was excluded that was already known by participants. As a result, 23 pictures were excluded from the 129 effective viewing pictures and were not considered in the final data analysis (N=106). Ultimately, a total of 16 effective serendipity cases were identified from 11 participants among the 106 pictures (see Table 6-3). Ten cases were identified after a successful drawing (positive emotion), while six cases were identified after a failed drawing (negative emotion). From another perspective, it was found that 14 out of 16 cases were identified with SCRs, while only 2 cases were identified without SCRs.

Table 6-3 Identified effective serendipity cases with marking scores over 4 on all dimensions

	Conditions	**Count**
Success	With SCRs	8
	No SCRs	2
Failure	With SCRs	6
	No SCRs	0
Total		16

During the interview session, participants also reported their reasons for their marks on each dimension. An interesting phenomenon revealed from this session was that all participants reported feelings of "unexpectedness" from the study in the following two aspects:

(1) The unexpected relationship between the pictures and the provided information:

> *"I never thought this picture would have such a relationship with me. I just thought the (provided) information would be some introduction of the Eiffel Tower [the provided picture]. But it turns out there would be a coming conference of my research area there! It really surprised me!"* (P16)

(2) The unexpectedness of the provided information content:

> *"It was really interesting that my supervisor had published this paper. I don't know it at all because it was published recently. I will go to read it later."* (P14)

Participants also expressed their opinions on the other three dimensions of "interesting", "relevant" and "beneficial" in relation to the provided information, such as:

"Interesting":

> *"My reading of the information mainly depends on my own interest in the information. I'll read information if it is interesting."* (P12)

> *"This is really interesting that your provided information is always out of my expectation, and I think this is a very good way for me to learn by this way [game] in my spare time!"* (P19)

"Relevant":

"Actually I was not so interested in the picture itself, but I was interested about the content it provided to me, because it was quite related to my research." (P22)

"When I read the information, I would first see whether the information is related to me. If it has relevance to me, for example, the information is related to my supervisor, I would have an interest in reading it, otherwise I may ignore it." (P2)

"Beneficial":

"I don't know about the (conference) information, and I think it is quite useful to me to learn about the recent chemistry conferences." (P12)

"I love such a way to receive beneficial academic information when I'm not so stressed. I can learn something new from it. I'm really glad about it." (P1)

There were two participants planning to attend some conferences who displayed great excitement about the given conference information and even wrote down the conference information after the study:

"I'm planning to attend some conferences soon and your conference information really gave me such a surprise!" (P22)

"I'm looking for a conference …… I'm really interested about this conference in Paris, and I'll go to check the detailed information right after the study!" (P9)

6.5.2 The impact of emotions on serendipity

Based on the results of the questionnaire, the impact of emotions was identified from two main aspects: 1) positive emotions and negative emotions during the game process; 2) discovered SCRs and no-SCRs during the text reading process.

6.5.2.1 Positive or negative emotions affect the scores on the questionnaire

Table 6-4 lists the four dimensions on the questionnaires, where marked by the participants under the two conditions of a successful drawing (positive emotion, N=63) and a failed drawing (negative emotion, N=41). The table shows the marking

differences between the two conditions. Participants tended to mark higher scores under positive emotions, especially for the dimensions of being "relevant" and "beneficial". The mean value of "relevant" for a successful drawing is 3.33 (SD=1.18), which means participants have mostly considered the provided text information to be somewhat relevant to him/her, while the mean value for a failed drawing is 2.37 (SD=1.24), which implies less relevance towards the provided information. The situation for the dimension of "beneficial", which refers to the value of the information provided to the participants, is quite similar to scores on the dimension of "relevant", with a mean value of 3.40 (SD=1.06) under positive emotions and 2.90 (SD=1.07) under negative emotions. A further *t*-test demonstrates that the differences are significant between the two groups for the dimensions of "relevant" ($t(102) = 4.009$, $p = 0.0001 < 0.05$) and "beneficial" ($t(102) = 2.324$, $p = 0.023 < 0.05$), with Cohen's *d* estimated at 0.8 and 0.5 respectively, which represents that there is an effect on the identified differences based on Cohen's (1992) guidelines. No significant differences exist on the dimensions of "unexpected" ($t(102) = 0.646$, $p = 0.52$) and "interesting" ($t(102) = 1.531$, $p = 0.129$).

Table 6-4 Marking values under different emotions

Condition	**Count**	**Unexpected**		**Interesting**		**Relevant**		**Beneficial**	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Positive emotions	63	3.27	1.33	3.59	0.93	3.33	1.18	3.40	1.06
Negative emotions	41	3.10	1.32	3.29	1.01	2.37	1.24	2.90	1.07

6.5.2.2 Discovered SCRs affect the marks on the questionnaire

By dividing the results from the questionnaires into the two groups of SCRs and no-SCRs, it is found that the mean score for the SCRs is obviously higher than that for the no-SCRs on all four dimensions (see Table 6-5). A *t*-test shows no significant differences for the dimensions of being "unexpected" ($t(102)=1.107$, $p=0.273$), "interesting" ($t(102)=1.683$, $p=0.095$) and "relevant" ($t(102)=0.817$, $p=0.416$), but there is significant difference for the dimension of being "beneficial" ($t(102)=2.324$, $p=0.023<0.05$), with Cohen's d estimated at 0.5, which represents that there is a medium effect on the identified difference between the two groups on this dimension. This result demonstrates that participants with emotional stimulation tended to mark higher than those without emotional stimulation, especially for the dimensions of "beneficial". Such information with SCRs is more likely to provide participants with the sense of serendipity.

Table 6-5 Marking values on different SCRs situations

Condition	**Count**	**Unexpected**		**Interesting**		**Relevant**		**Beneficial**	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
SCRs	66	3.32	1.19	3.59	0.91	3.03	1.28	3.37	1.02
no-SCRs	38	3.00	1.52	3.26	1.03	2.82	1.31	2.89	1.13

6.5.2.3 Multiple comparisons

Following a further comparison of the four groups (i.e. positive emotions with SCRs—Group 1, positive emotions with no-SCRs—Group 2, negative emotions with SCRs—Group 3, and negative emotions with no-SCRs—Group 4), it can be found that participants in Group 4 marked lowest in all dimensions across the four groups, with a mean value of 2.67 (SD=1.32) on the dimension of "unexpectedness", 2.73

(SD=0.96) on the dimension of "interesting", 2.13 (SD=1.25) on the dimension of "relevant" and 2.40 (SD=1.06) on the dimension of "beneficial" (Table 6-6).

ANOVA analysis shows that no significant differences were discovered on the dimension of "unexpectedness" among all four groups ($F(3,100)=0.993$, $p=0.399$), but there are significant differences among the four groups on the dimensions of "interesting" ($F(3,100)=3.696$, $p=0.014$, $\eta^2= 0.1$), "relevant" ($F(3,100)=5.641$, $p=0.001$, $\eta^2= 0.145$) and "beneficial" ($F(3,100)=4.094$, $p=0.009$, $\eta^2= 0.109$). Table 6-7 shows post-hoc tests on the value of p for the three dimensions among the four groups, according to the table, it can be found that:

- On dimensions of "interesting" and "beneficial", Group 4 is significantly different from the other three groups;
- On dimension of "relevant", Group 4 is statistically different from Group 1 (p = 0.001) and Group 2 (p = 0.006); while Group 3 is statistically different from Group 1 (p = 0.005) and Group 2 (p = 0.03).

The results from the ANOVA implies that the cases in Group 4 —— negative emotion with no-SCRs —— is probably with the least possibility for participants to experience serendipity when compared to the other three groups.

Table 6-6 ANOVA of the questionnaire marking

Conditions		**Count**	**Questionnaire Result**							
			Unexpected		Interesting		Relevant		Beneficial	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD
Positive	With SCRs (Group 1)	40	3.30	1.22	3.58	0.93	3.37	1.20	3.50	1.03
Emotions	no-SCRs (Group 2)	23	3.22	1.54	3.61	0.94	3.26	1.18	3.21	1.09
Negative	With SCRs (Group 3)	26	3.35	1.16	3.62	0.90	2.50	1.24	3.19	0.98
Emotions	no-SCRs (Group 4)	15	2.67	1.32	2.73	0.96	2.13	1.25	2.40	1.06

Table 6-7 Post-hoc test of *p* values among different groups on three dimensions

Groups	Interesting	Relevant	Beneficial
Group 1-Group 2	0.89	0.719	0.3
Group 1- Group 3	0.863	0.005*	0.242
Group 1- Group 4	0.003*	0.001*	0.001*
Group 2- Group 3	0.98	0.03*	0.933
Group 2- Group 4	0.005*	0.006*	0.019*
Group 3- Group 4	0.004*	0.352	0.02*

*the main difference is significant at the 0.05 level

6.5.3 Comparisons among the three designed conditions of algorithms

Figure 6-8 demonstrated participants' ratings of the questionnaires under the three different designed conditions. Only the marks of four or five are considered to be effective values on the corresponding dimension.

An overview of the figure, it can be found that ratings on Condition 1 and Condition 2, the two conditions that are designed following the developed serendipitous algorithm, performs much better than Condition 3, which is a conventional form of "picture-and-information". On dimensions of "unexpectedness", "interesting" and "relevant", the Condition 1 achieves the most effective scores from participants, which shows that this type of information impact most on participants on the three dimensions. The Condition 2 performs best on the dimension of "beneficial", implying that the provided related conference information can bring most valuable information to participants comparing to the other two. The results provide evidence that the designed algorithm is with the functionality to help facilitate participant's serendipitous encountering.

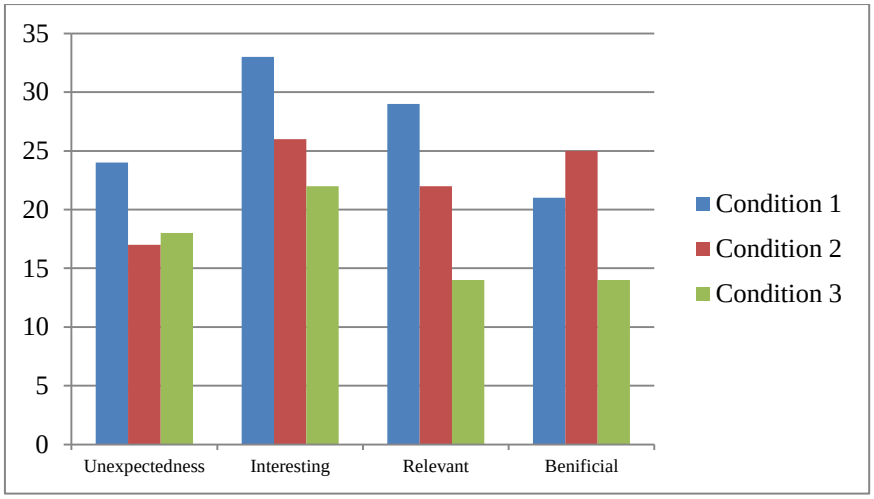


Figure 6-8 Participants' effective ratings under three different conditions

6.6 Discussion

6.6.1 The role of emotions in serendipitous encountering

The study has found that the emotional valence (positive vs. negative) has influenced participants' marking on the dimensions of "relevant" and "beneficial", while the SCRs experienced have influenced marking on the dimension of "beneficial". No significant difference was discovered on the dimension of "unexpectedness" under both situations. A possible explanation may attributable to the designed algorithm and the "Wizard of Oz" approach. In the interview session, 21 out of 26 participants reported that an important reason for their marking on this dimension (with a score of 4 or 5) was the unexpected relationship between the provided picture and the text information. They considered that all the information would be exactly the same as the controlled condition (condition 3), and both the socialisation-based information (condition 1) and the research background-based information (condition 2) was not something they had expected. However, such unexpectedness could decrease as the study continued. As each participant needed to

draw six pictures in two rounds of the three conditions, at the second time of reading the information, it may not have been as unexpected as the first time. This may also explain why although a majority of participants reported they found the information to be unexpected, the mean scores for this dimension were around 3 under all different emotional situations.

For the dimensions of "relevant" and "beneficial", participants under positive emotions marked significantly higher than those under negative emotions. This implies that participants were more concentrated on the provided information under positive emotions. Research findings by Fulton (2009) and Savolaien (2014) show that a positive emotion is a prime motivator for beginning an information seeking activity, while psychology studies have also proved that a positive emotion can broaden the scope of people's attention and thinking (Fredrickson and Branigan, 2005). These features of positive emotions would probably prompt a participant's motivation to read and think about the information more actively and thus make it more likely that they would find the value of the provided information. Current serendipity studies have also highlighted how the "act of noticing" is an important factor in the process of making connections (Rubin et al., 2011; McCay-Peet and Toms, 2015), which is a key process in connecting any encountered information with one's previous experience or knowledge before finally leading to serendipity. The broadened scope of attention brought by a positive emotion will probably contribute to an individual's ability to notice. A typical example of this explanation was identified by one participant (P7) during the interview session, when he argued that if he had drawn the picture incorrectly, he would keep thinking about his mistakes during the drawing session even though he was reading the text information. Thus, his

attention was not strictly focused on the provided information and this certainly affected his perceptions and connection-making while dealing with the reading information.

For the dimensions of “beneficial”, participants marked significantly higher in cases with SCRs than in those with no-SCRs. The results reflect the effectiveness of employing EDA sensors in capturing participants’ emotional responses. Although no significant differences are identified on the dimension of “interesting”, it should still be noted that there is a difference between “interesting” and “relevant”. Under both emotional states (i.e. emotion valence and SCRs), participants’ marking results are not consist for these two dimensions. For example, many participants in the study marked the text information in condition 3 as “interesting”, but not “relevant”. This finding is in accordance with Pontis et al.’s (2015) study, in which they argued that participants may also find unacknowledged information as “interesting”, while “relevant” requires a more cognitive and situational level, which is correlated to one’s background knowledge and recent focuses. This study demonstrates that it is better to separate “interesting” and “relevant” to better evaluate serendipity.

6.6.2 Research methods applied in the study

This study is also a new attempt to combine other HCI research methods into the study of serendipity. Both Eye-tracker and EDA sensors were employed in the study through the game application, and the equipment helped researchers to capture substantial objective data.

I believe the employment of such ideas will open new windows for other researchers to combine more HCI research methods into the research of serendipity. However, although the majority of participants (N=23) reported they were totally in

favour of the applied research methodology, and even showed great interest in their collected data, there are still some issues worthy of further discussion.

First, it is worth highlighting the employment of the thinking-aloud session. The collected objective data from the EDA sensors and the Eye-tracker both helped us to understand each participant's emotional experience during the study, and played a vital role in helping participants to recollect their memories during the study. Traditional research methods on serendipity, such as the widely used methods of interviews and critical incidents, have mainly been based on participants' own memories, which makes serendipity rather a subjective phenomenon and there are always challenges to understand the instant context in which serendipity occurs. By presenting participants with the captured data during the thinking-aloud session, participants were better able to describe their experiences and instant opinions of some certain episodes, such as the discovered peaks on the SCL data. I believe such a research method provides possible solutions for researchers to understand the instant context in which an individual experiences a serendipitous encounter.

Bogers et al. (2013) pointed out that it is beneficial to 'keep controlled experiments designed to measure serendipity as natural as possible'. What discovered from the study is that there is a necessity to pay close attention to participants' initial status before they begin the study, because it is better to design research measures to help them find a more natural status, especially in controlled laboratory settings. Nearly half of the participants (N=13) highlighted the importance of the time for relaxation before their sketching. The original aim of the study for this stage was to collect each participant's baseline SCL signals, and it was not expected in the beginning that such a process could help participants to reduce their stress levels and

to calm down in the laboratory setting. One participant (P22) reported that she was really experiencing a high level of stress when she entered the laboratory; however, this was largely reduced during the five-minute relaxation period with her eyes closed. Special consideration needs to be taken of this point when conducting similar studies in the future, especially in controlled research settings.

It is also apparent from the study that a leisurely and relaxed context, such as providing a game application to participants, can effectively help them to reduce pressure in a controlled study setting. Although the relaxation process largely helped participants to reduce their stress levels, it was inevitable that some stress remained when participants set about completing their tasks. An obvious reflection of this phenomenon is that the collected SCL value was always higher during the first picture drawing process, when compared to the later pictures. During the thinking-aloud sessions in the interviews, several participants reported that although they were still nervous at the beginning, they were quickly attracted by the game and this removed the pressure. I believe a relatively relaxed context, which can quickly immerse participants, is another effective condition for helping them to conduct studies in a more natural state.

However, regardless of the efforts on preparation of the study, there was still one participant (P3) who argued that he would prefer to finish such a study without any monitoring. This argument conforms to Bogers et al.'s study (2013), which found that participants who were not being monitored were more likely to experience serendipity. This finding raises implications for the design of studies, confirming that additional possible solutions should be explored in any future research to avoid

monitoring participants and to make the research environment "as natural as possible".

6.7 Conclusion

Although the study of serendipity continues to garner an increasing amount of research activity, the role of emotions is often neglected in existing studies. This chapter contributes to this research gap with the findings that emotions did play a role during participants' experiencing of serendipity. Through the "Wizard of Oz" approach, which is based on the developed serendipitous algorithm, this study presented a laboratory research setting where participants were able to experience serendipity. With the help of the subjective data collected from questionnaires and interviews, in addition to the objective data collected from the Eye-tracker and EDA sensors, the study has found that a positive emotion (success in drawing) and/or experiencing an emotional stimulus (with SCRs) were conducive to participants' scoring high marks on the evaluated dimensions of serendipity, which implies there is an increased possibility for participants to encounter serendipity under such conditions.

The results from the study bring new design implications that extend beyond current design strategies for serendipity. It is clear that attention should be paid to emotional designs, especially those that can lead to positive emotions and/or can stimulate an individual's emotional responses. In addition, the successful of the study also provides evidence for the developed serendipitous algorithm in facilitating serendipity, and rather than applying the "Wizard of Oz" approach, the following chapter will introduce the details of the proposed algorithm and how to employ the

algorithm to real data setting to further evaluate its effectiveness in recommender systems.

CHAPTER 7

SERENDIPITOUS RECOMMENDER SYSTEM DEVELOPMENT AND EVALUATION

Research questions addressed in this chapter:

How to design information technology such as recommender systems to facilitate Chinese scholar's serendipitous encountering?

[1] How to develop serendipitous algorithm in recommender systems based on the understanding of serendipity from information research?

[2] How to implement the algorithm in real data setting?

[3] How to evaluate the effectiveness of the developed recommender system?

7.1 Introduction

This chapter explores how to facilitate Chinese scholar's serendipitous encountering through existing information technology. According to Makri et al.'s (2014) review, existing design implementations in supporting serendipity and can be divided into two main categories: information visualisation and recommender systems. While information visualisation is more of a design technique that to present data via visual layouts or visual interaction methods to help individual's cognition of the data (Card et al., 1999; Liu, Cui, Wu, and Liu, 2014; Nazemi, Burkhardt, Hoppe, Nazemi, and Kohlhammer, 2015), recommender systems have paid particular interest in the

topic of "serendipity" in recent years (e.g. Yamaba, Tanoue, Takatsuka, Okazaki, and Tomita, 2013; Sun, Zhang, and Mei, 2013; Adamopoulos and Tuzhilin, 2015; de Gemmis, Lops, Semeraro, and Musto, 2015). The literature review session in Chapter 2 has demonstrated that researchers in this area have used both content-based filtering and collaborative filtering to design recommender systems that can provide serendipitous recommendations to users. However, as Makri et al. (2014) argued, most of current designs and implementations on serendipity "are based on an intuitive (rather than empirically grounded) understanding of the phenomenon" (p.2183). These developed recommender systems are mainly dependent on the designer's intuitive perception on serendipity and lack empirical studies to look into how users experience serendipity, which, however, is the main research point by information researchers.

Chapter 5 investigated how Chinese scholars experience serendipity and a context-based model was constructed. The result showed that similar to exist serendipitous encountering of Westerners, Chinese scholars also experience serendipity with three main processes: any unforeseen means and/or unexpected content of information cause *unexpectedness* to a participant, then a *connection is made* between the unexpectedness and a participant's precipitating condition (i.e. visceral need, conscious need and previous knowledge/experience), and once a substantial *value* (e.g. find answers to previous information needs) or an emotion *value* is recognised by the participant, serendipity occurs. The identification of the three processes (i.e. encountering unexpectedness, connection-making and recognising the value) from the diary-based study depicted a clear picture of how Chinese scholars experience serendipity.

In Chapter 6, a controlled laboratory research setting was designed to investigate the role of emotion played in serendipitous encountering. The idea of designing the three conditions through the Wizard of Oz approach was based on a new information theory-based serendipitous algorithm, and the result from the study showed the effectiveness of such design in facilitating serendipity.

In this chapter, a systematic introduction of the algorithm is given in Section 7.2, and Section 7.3 introduces how the algorithm is implemented in real data setting. Section 7.4 evaluates the effectiveness of the algorithm through a vitro experiment in the cyberspace and a case study on real users. Section 7.5 gives the discussion and conclusion of this chapter.

7.2 Proposed Serendipitous Algorithm in Recommendation Setting

There are two major concerns in providing serendipitous encountering in the recommendation system design: the first concern is how to balance "unexpectedness" and "useful". As pointed out by Adamopoulos and Tuzhilin, (2011), there should be "a most preferred distance" between the two values, as the high level of unexpectedness may cause user's dissatisfaction of the recommended information, while users may also lose interest to that information with a low unexpectedness. The second concern is how to combine "insight" into system design to stimulate the process of "making connections".

The two concerns are addressed from the following perspective of "relevance" with two hypotheses:

- Hypothesis 1: Given the information that is highly relevant to a user's personal profile, the information would also of a high potential value to the user;

- Hypothesis 2: A user will be unexpected to the information that is relevant to his profile while is not previous acknowledged or known by the user.

Consider a target user A, who is the user that will be provided with the recommended information, a user B who is highly relevant to user A and a user C who is highly relevant to user B while is not known by user A. The user A may experience serendipity by providing the information of user C, which is unexpected to him/her, and by providing the relationship between user B and user C, which may further cause interestingness or usefulness to user A. The following part of this section illustrates a detailed implementation of the algorithm.

1. Target user

Consider a table of a target user profile U_1 with a category set $C= \{C_1, C_2, C_3 \ldots C_i \ldots C_n\}$, where C_i represents the i-th category of the user profile. All the categories are arranged through the value of their weights in the user profile. The weight can either be a given weight by the dataset or calculated through clustering analysis (Rohfl, 1992). In order to simplify the introduction of the proposed algorithm here, it is more convenient to set the weight for each C_i which is given by the dataset in the very beginning. The weight of C_i is larger than C_j ($i>j$) in C set:

$$w_c = \{w_{C_1}, w_{C_2}, \ldots, w_{C_i}, \ldots, w_{C_i}, \ldots, w_{C_n} \mid w_{C_i} \geq w_{C_j}, i > j\} \quad (1)$$

For each category set C_i, consider $C_i= \{a_1, a_2, a_3 \ldots a_i \ldots a_n\}$, where a_i is the corresponded attribute to each vector C_i. In particular, for each a_i represents the dimension according to which a new user profile may be produced (i.e. author of literatures; musicians). The values for each a_i are also arranged by their weight in each vector C_i

and can be calculated through semantic analysis such as the *tf*idf* weight (term-frequency times inverse document frequency) calculation (Pazzani and Billsus, 2007):

$$w(t,d) = \frac{tf_{t,d} \log\left(\frac{N}{df_t}\right)}{\sqrt{\sum_i \left(tf_{t_i,d}\right)^2 \log\left(\frac{N}{df_{t_i}}\right)^2}} \tag{2}$$

Where *w(t,d)* represents for the weight of a term *t* in a document *d*, and it is a function of the frequency of *t* in the document *(tft,d)*, the number of documents that contain the term *(dft)* and the number of documents in the collection *(N).* As a result, the weight for a category set Ci is determined by the weight of each attribute in the set:

$$w_{c_i} = \{w_{a_1}, w_{a_2}, \ldots, w_{a_i}, \ldots, w_{a_j}, \ldots, w_{a_n} \,|\, w_{ai} > w_{a_j}, i > j\} \tag{3}$$

2. Screen the weight

As been pre-defined that C_1 with the largest weight in the C set and a_1 with the largest weight in the C_i set. Set a threshold τ to eliminate the low weight value from the user profile U_1:

$$w_{c_i} = \{w_{a_1}, w_{a_2}, \ldots, w_{a_i}, \ldots, w_{a_j}, \ldots, w_{a_n} \,|\, w_{ai} > w_{a_j}, i > j\} \tag{4}$$

Similarly, set a threshold θ to eliminate the low weight value from the C_i set:

$$w_{C_i} = \{w_{C_i,a_1}, w_{C_i,a_2}, w_{C_i,a_3}, \ldots, w_{C_i,a_i} \,|\, w_{Ci,a_i} \geq \theta\} \tag{5}$$

3. Generate a new user profile

A new user profile U_{i+1} is produced according to each a_i in the C_i set. Here, the generation of the user profile arranges from the largest weight of w_{C_i,a_1} to the smallest weight of w_{C_i,a_i} .

4. Iteration and End condition

Based on the weight arrangement in a user profile, it is intuitional that for an attribute a_i with a large weight, it is also with more possibility for the current user to have acknowledged about the information of a_i. In other words, the probability for a current user U_i to make connection with the next user profile U_{i+1} is proportional to the weight of the attribute in current user profile:

$$P(U_{i+1}|U_i) = \lambda w_{c_i} * w_{c_i,a_i} \tag{6}$$

where λ is the proportionality coefficient of the probability to the relevant weight.

The probability of making connections by target user U_1 to *i*-th user can be further extended if only the generated user is always new to the prior generated ones:

$$P(U_i|U_1) = P(U_2|U_1) * P(U_3|U_2) * \ldots * P(U_i|U_{i-1}) \tag{7}$$

The iteration to find the next user would not continue until it meets the following two end conditions:

- the generated user is no longer new to all the previous generated users;
- $P(U_i|U_1)$ comes to a threshold δ, where δ represents an appropriate threshold of the probability.

The reason to set the threshold δ is to ensure the effectiveness of the iteration process. This is because if $P(U_i|U_1)$ comes too large, the recommended information may fail to bring the target user with the sense of unexpectedness, as the recommendation may probably have been acknowledged by the user; however, if the value of $P(U_i|U_1)$ is too small, the recommended information may be too irrelevant to the target user and he/she may lose interest on it. Hence the setting of the threshold δ is a very

important step for the iteration process and it needs to be further identified based on empirical studies in the future. Once the recommendation list is generated within the threshold δ, they can be recommended to the target user by selecting the item with the highest values of $P(U_i|U_1)$.

5. Recommendation

When the iteration is finished, the content with the largest weighted category in cur-rent candidate will be provided to the target user, in addition with the relevant information of the previous searched users that result in the current user.

6. An example of the proposed algorithm

An example of the proposed algorithm is provided in Figure 7-1. Consider *Ann* as the target user (U_1) with different literature categories of *{A, B, C}* in her personal library, whose weight is *{0.5, 0.3, 0.2}* (Figure 7-1-a). The author names of the literatures are set as the attributes for each category and according to the *tf*idf* weight calculation, there are three values $\{a_1, a_2, a_3\}$ in category A with the weight $W'A = \{0.6, 0.3, 0.1\}$. Set $\lambda=1$ for each probability of the current user to find the next user profile, the probability for Ann to find *a1*'s profile (U_2) can be calculated according to equation (6):

$$P(U_2|U_1) = w_A * w_{A,a_1} = 0.5 * 0.6 = 0.3 \quad (8)$$

The profile of *a1* is then produced as Figure 7-1-b. Likewise, among the four authors in the *D* category, author *d1* (U_3) weights largest and then produce *d1*'s profile (Figure 7-1-c):

$$P(U_3|U_2) = w_D * w_{D,d_1} = 0.4 * 0.5 = 0.2 \quad (9)$$

According to equation (7), the probability for *Ann* (U_1) to find *d1*'s profile (U_3) is:

$$P(U_3|U_1) = P(U_2|U_1) * P(U_3|U_2) = 0.3 * 0.2 = 0.06 \quad (10)$$

Set the threshold δ as 0.06, then the iteration of the algorithm stops and recommend literatures of category *F* in *d1*'s profile to *Ann*, in addition with the relevant information of *d1* and *a1*. For example, the recommended information can be "these papers (category *F*) are most stored by *d1*, who had published papers *(d1, d2, d3, d4)* with *a1* before".

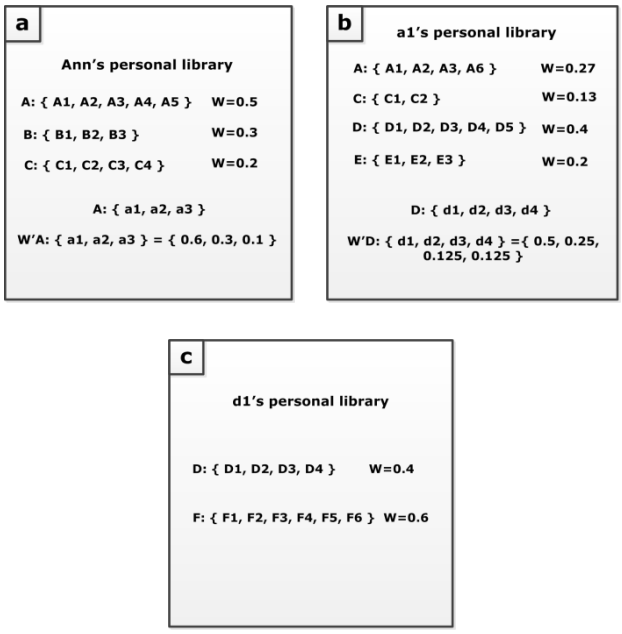


Figure 7-1 An example of the proposed algorithm: (a) target user *Ann*'s personal library; (b) user *a1*'s personal library generated by *Ann*; (c) user *d1*'s personal library generated by *a1*

7. Description of the Proposed Algorithm

The proposed algorithm is collaborative filtering based, hence it is more appropriate to those dataset whose content is generated by different users, according to which the next user's profile will be easier to produce for a current user. The algorithm relates with serendipity from the following three aspects:

— *Unexpectedness*: by setting the value of probability. In an identified threshold δ, the unexpectedness of the information to a target user is inversely related to the

magnitude of probability. The smaller probability for a target user to find an-other user, the more unexpectedness he/she receives from the provided information of the current candidate.

– *Connection-making*: by providing the information of the searched clues which demonstrates the relationship between the provided user (recommendation source) and the target user. As aforementioned that the ability to connect the new clue with previous knowledge/experience is a key element in the occurrence of serendipity, and thus there is a necessity for the designers to provide the design clues which can contribute to a customer's noticeability or attention to connect the provided information with his/her personal profile. In the provided example of Figure 7-1, such process of connection-making is constructed by showing the relationship between *d1* and the target user, who had published paper together before.

– *Value*: by generating the next user's file according to the weight arrangement of the attributes; those with larger weights are considered as priorities. This is because the larger weight the attribute is, the more possibility it may have to satisfy the target user's need/concern, and finally brings more potential value to the user.

7.3 Implementing the Algorithm in Real Data Setting

This chapter aims to evaluate the effectiveness of the algorithm in real data setting. The Movielens data set is selected to implement the algorithm, this is because it is a data set that widely used by researchers in recommender systems to test their developed algorithms, including those serendipitous ones (e.g. Adamopoulos and Tuzhilin , 2011; de Gemmis et al., 2015).

The HETREC2011-MOVIELENS-2K dataset was selected, and it was available at the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems, HetRec 2011 (Cantador, Brusilovsky, and Kuflik, 2011). The original dataset contains 855,598 rating assignments on a 10-point Likert scale from 0.5 to 5.0 (step 0.5), provided by 2113 users on 10,197 movies (sparsity 96.03%). In the dataset, the movies are linked to the Internet Movie Database (IMDb) and RottenTomatoes (RT) movie review systems. Each movie has its IMDb and RT identifiers, English and Spanish titles, picture URLs, genres, directors, actors, audience' and experts' ratings and scores, countries, and filming locations. It also contains the tag assignments of the movies provided by each user.

7.3.1 A general description of the algorithm

Figure 7-2 gives the abstract process of implementing the algorithm in this data set, where User A is considered as the target user.

1. Choosing the most related director in the user profile as the start point of recommendation. This step is corresponding to the first two steps of the algorithm. "Directors" of movies are considered as an important attribute in the target user's profile, and the "most related" director is calculated through two parts of data: the number of movies watched by the user with respect to the director, and users' effective ratings on these movies. Only those ratings no less than 4.0 were considered as effective ratings. As a result, the director with the most effective ratings from the user was screened as the most related director (Director X in Figure 7-2).

2. Generate the next user profile. Once the director is chosen, another user profile is generated which covers all the users in the dataset that have watched the movies of

this director. This profile is arranged by the number of effective scores on the director rated by different users, which is calculated similarly with step 1. The user with the largest effective score on the director throughout the whole dataset is selected as the next user (User B in Figure 7-2).

3. Iteration and find the end user for recommendation. This step is corresponding to step 4 of the algorithm. The Pearson similarity (Resnick et al., 1994) is used to calculate the distance between the target user (User A) and the generated user in the last step (User B), and only when the calculated Pearson value falls into a determined threshold δ that the second user will be selected as the end user to generate recommendations, or else the algorithm will iterate the calculation of the last two steps from the second user, and to continue generating the next user. It should be noted that the Pearson similarity is always compared between the target user (User A) and the new generated user during the iteration process. A larger value of Pearson similarity represents a larger possibility for the active user to find the selected user.

4. Recommendation. Once the end user is selected, the item list in the user profile will be re-arranged based on the measurements from Murakami et al. (2007) and Ge et al. (2010). The system then recommends the new generated items on the recommendation list to the active user.

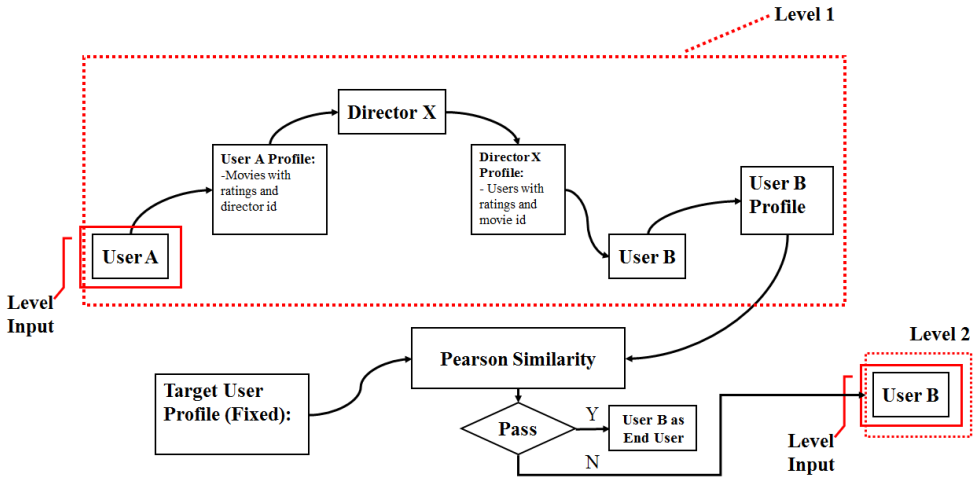


Figure 7-2 A general picture of the algorithm

7.3.2 Three main blocks to implement the algorithm

Three main blocks were developed during the implementing of the algorithm, which corresponds to the three parts of serendipity: the connection-making block, the unexpectedness block and the usefulness block.

(1) Block 1: Connection-making block

This block mainly addresses the problem of finding the most relevant director and generating new user profile that is relevant to the target user. Figure 7-3 shows the develop logic for a target user A to find the active user B who will generate the recommendation lists through the measurement of choosing the most related director.

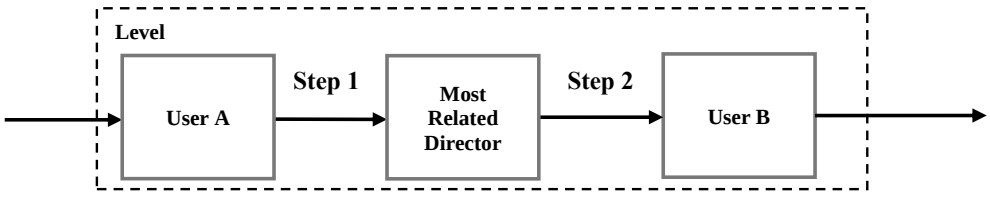


Figure 7-3 The develop logic of from target user A to find active user B

There are two steps involved in Figure 7-3: the first step is to find the most related director according to the target user A; and the second step to find user B according to the selected director's profile. The screen process for each step is similar, and two parts of data were taken into consideration: number of movies with effective ratings and user's effective ratings sums on the movies. Considering the first step, based on the target user's profile, which is a rating log list (see Figure 7-4 as an example). The program starts by selecting a rating log in the target user's profile, and then to judge whether the rating is an effective rating. Only those ratings no less than 4.0 can be considered as effective. An effective rating list can then be generated, and those logs with the same director name can be merged, and finally group into a list which contains different groups of directors with effective rating numbers and the sums of these ratings. It should be noted that in the following steps, the number of movies with effective ratings is considered as the prior condition, and the group with largest effective rating numbers is selected. As a result, the correlated director in this group is considered as the most related director with respect to the target user A. Take the example of Figure 7-4 to explain the process of choosing the director: in the profile of a target user A, there are eight movies directed by five directors of J, K, L, M and N. Only those ratings no less than 4.0 can be considered as effective ratings. Therefore, the Movie ID 5 was excluded during the grouping process, and after grouping, the User A gave 3 effective ratings to the movies directed by J. The director J was finally selected as the most related director with respect to the target user A.

User A – Before Group		
Movie ID	Director ID	Rating
1	J	4.5
2	K	4.0
3	J	5.0
4	J	4.5
5	K	3.0
6	L	4.5
7	M	4.5
8	N	5.0

Effective:
Rating >= 4.0

User A – After Group		
Director ID	Effective Rating Number	Effective Rating Sum
J	3	14.0
K	1	4.0
L	1	4.5
M	1	4.5
N	1	5.0

Figure 7-4 An example of the process for selecting most related director

However, during the real data processing, a problem was revealed in selecting the most related director, which is a situation that there is more than one most related director. Take Figure 7-5 as example, after the grouping, the target user F had two directors of L and M with the same largest numbers and sums of effective rating. In this case, both directors were selected as the most related directors. Figure 7-6 demonstrated the workflow for selecting the most related director(s) from a target user, and the final output can either be a single director or a group of directors.

User F – Before Group		
Movie ID	Director ID	Rating
2	K	4.0
5	K	3.0
6	L	4.5
7	M	4.5

Effective:
Rating >= 4.0

User F – After Group		
Director ID	Effective Rating Number	Effective Rating Sum
K	1	4.0
L	1	4.5
M	1	4.5

Figure 7-5 An example of the situation with more than one most related directors

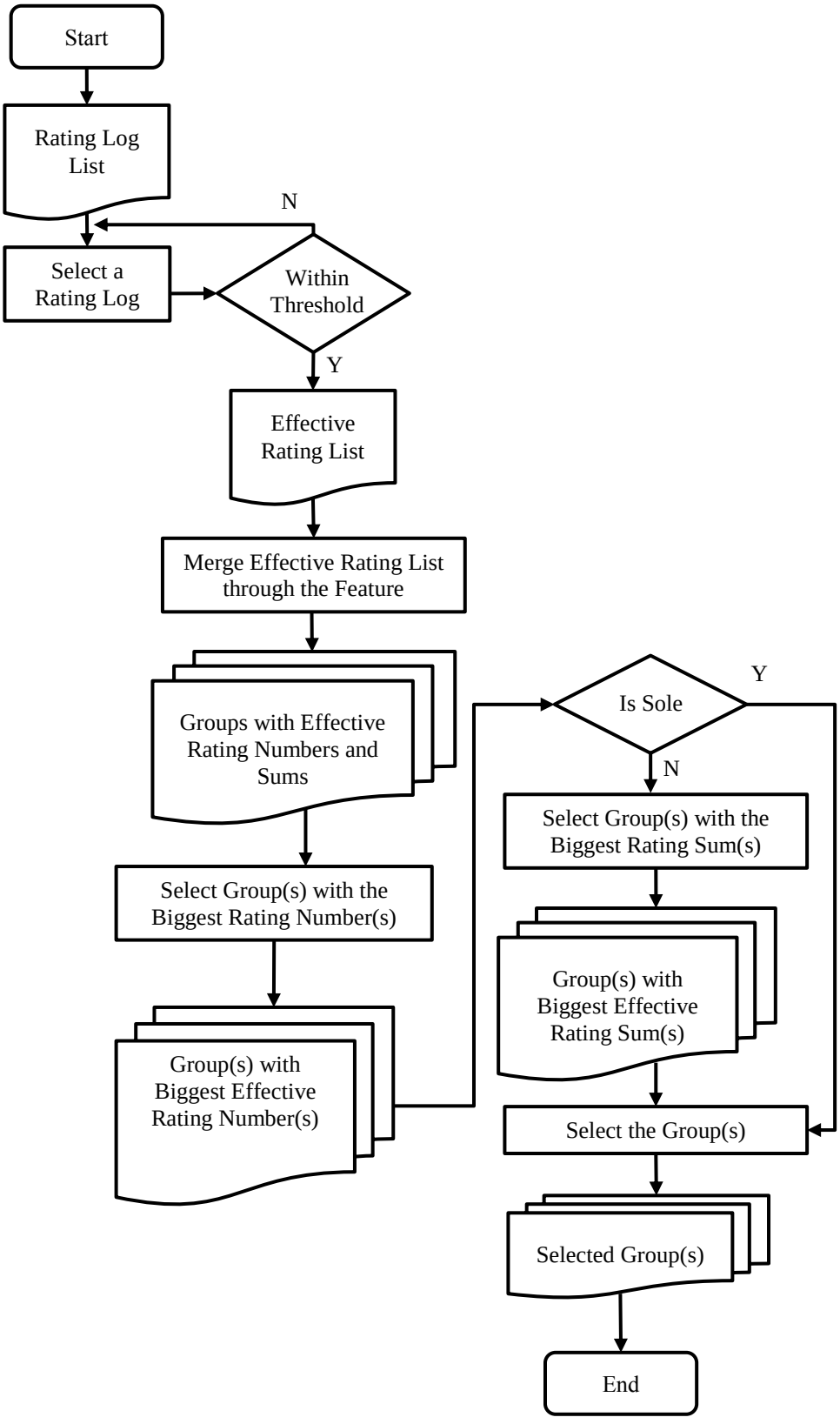


Figure 7-6 Workflow of selecting the most related director based on the target user's profile

Once the most related director is found, the program continues to the second step, which is to generate new user profile for recommendation. The generated user has a relationship with the director: the user gives the most effective ratings to the director. For example, in Figure 7-7, both the profiles of user C and user D were generated due to the most related directors. User C is selected because he/she gave the most effective ratings to director 1, and user D gave the most effective ratings to director 2. Here, for convenience, User C and User D can be called as the "user fans" of director 1 and director 2 respectively. The way to find the user fan of a director is the same to the first step, which also follows the workflow in Figure 7-6. The start input is the selected director's profile from the first step and the output is the most related user fan to this director. Figure 7-8 shows the algorithm for the target user to find the active user involved in Figure 7-3.

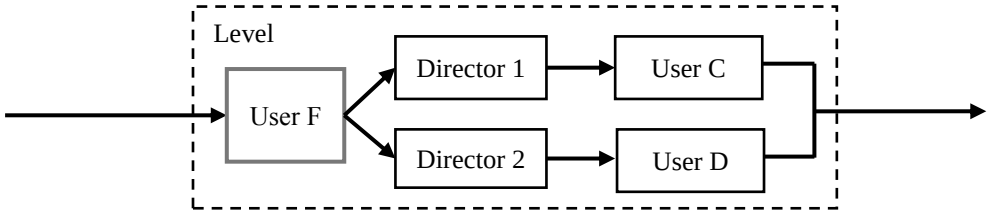


Figure 7-7 The situation with more than on most related directors

Algorithm: Connection-making

Input: A target user;

Output: An active user or An active users' set;

L : Input user's rating logs;

$Thre$: The threshold to determine if this rating log is effective, here it's 4.0;

$RNum_e$: A list of the number of effective ratings based on different directors;

$RSum_e$: A list of the sum of effective ratings based on different directors;

D : The reference director;

U_D : User list from the reference director D except the input user;

for *each rating log* x *in* L **do**

if $x \geq Thre$ **then**

Add x into the effective log set L_e ;

end

end

Merge the logs in L_e based on the directors and update them into $RNum_e$ and $RSum_e$;

Find the maximum value N in $RNum_e$;

if N *is sole in* $RNum_e$ **then**

Set the director with N into the reference director D ;

end else

Remove all the movies without the value of rating number N in $RSum_e$;

Find the maximum value S in $RSum_e$;

if S *is sole in* $RSum_e$ **then**

Set the director with S into the reference director D ;

end else

Set the all the directors into the reference director set D_{refer} ;

return a set of active users most related to the directors in D_{refer} respectively ;

end

end

return the active user most related to the director D ;

Figure 7-8 Algorithm for the target user to find active user

A similarity calculation was then conducted between the selected user fans (User C and User D) and the target user F. If the calculated similarity is within the defined threshold δ, this user was then selected as the active user to generate recommendations, or else the program will continue the iteration to find other active users. For example, in Figure 7-7, there are three different conditions:

- Condition 1: one of the similarities between the selected users and target user is within the threshold, while the other is not within the threshold. For example, if the similarity between user C and the target user F (Sim(C,F)) is within the threshold, and the similarity between user D and the target user F (Sim (D,F)) is not within the threshold, then the program will stops here and selects user C as the active user to generate recommendations.
- Condition 2: both Sim (C,F) and Sim (D,F) are not within the threshold, then the iteration will continues. The user C and user D will be considered as new input in Figure 7-6 to continue finding different most related directors based on these two users, and the iteration will only stop when at least one selected user (e.g. user X) whose similarity with the target user F (Sim (X, F)) falls into the threshold.
- Condition 3: both Sim (C,F) and Sim (D,F) fall into the threshold. In this case, the program will randomly pick one of the two selected users as the end active user and generate recommendations.

Here, Pearson-correlation (Resnick, et al., 1994) is used to calculate the similarity. It is popularly used in user-based algorithms to measure the extent of similarity between users *u* and *v*:

$$w_{u,v} = \frac{\sum_{i \in I}(r_{u,i} - \overline{r_u})(r_{v,i} - \overline{r_v})}{\sqrt{\sum_{i \in I}(r_{u,i} - \overline{r_u})^2}\sqrt{\sum_{i \in I}(r_{v,i} - \overline{r_v})^2}} \quad (11)$$

where the $i \in I$ summations are over the items that both the users *u* and *v* have rated, $r_{u,i}$ is the rating of *u*th user on the *i*th item and $\overline{r_u}$ is the average rating of the co-rated items of the *u*th user. Conventionally, the larger of the similarity between a user *v* and the target user *u*, the larger probability will be for user *u* to find user *v*.

The whole architecture for running the connection-making blo ck is introduced in Figure 7-9, and Figure 7-10 gives the pseudo-code for this block.

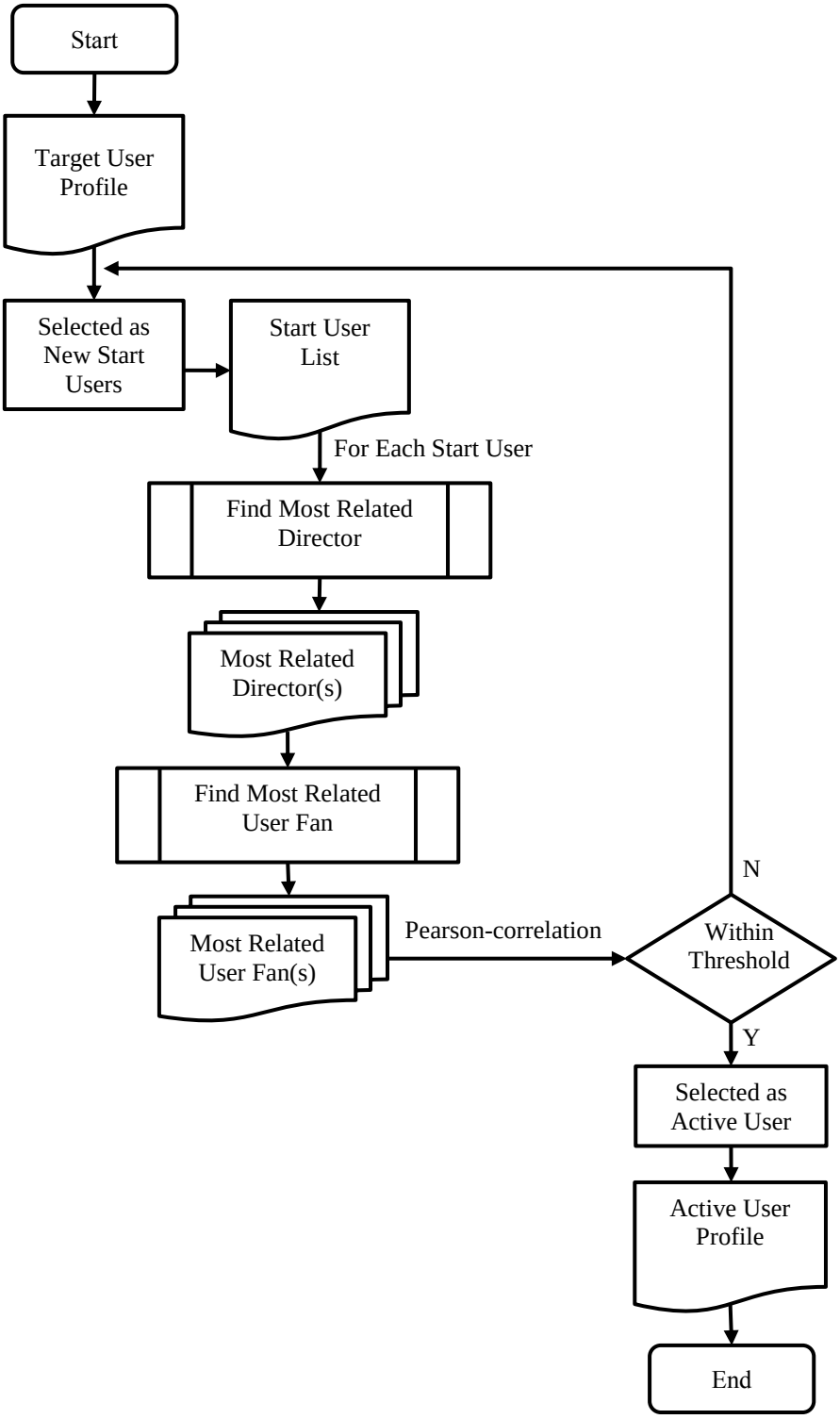


Figure 7-9 The architecture for the whole “connection-making block”

Algorithm: Connection-making integration

Input: The user set U_{input}**;**

Output: An movie list M_{output}**;**

U_{active}: The user set which contains all the active users;

$Sim(u_A, u_B)$: A function to calculate the similarity between user A and B;

$Thershold$: The standard to determine if "connection-making" should be ended;

while U_{input} *is not empty* **do**

for each user u in U_{input} **do**

Deploy "Connection-making algorithm" by inputting u *;*

Form the active users' set U_{active} ;

for each user u' *in* U_{active} **do**

if $Sim(u_{target}, u') < Thershold$ **then**

Set the user u' as the selected user $u_{selected}$;

return the movie list M_{output} from the selected user $u_{selected}$ for output;

end

end

end

Update U_{input} with U_{active};

end

Figure 7-10 The pseudo-code for the whole "connection-making" block

(2) Block 2: Unexpectedness block

This part of work is conducted from the following two perspectives: the first is to expand the movie list in the target user's profile, and generate an "expected movie

list". The second part is to filter active user's profile, which was generated from the first "connection-making" block, by the expected movie list and finally obtain an "unexpected recommendation list". The architecture for this block is introduced in Figure 7-11.

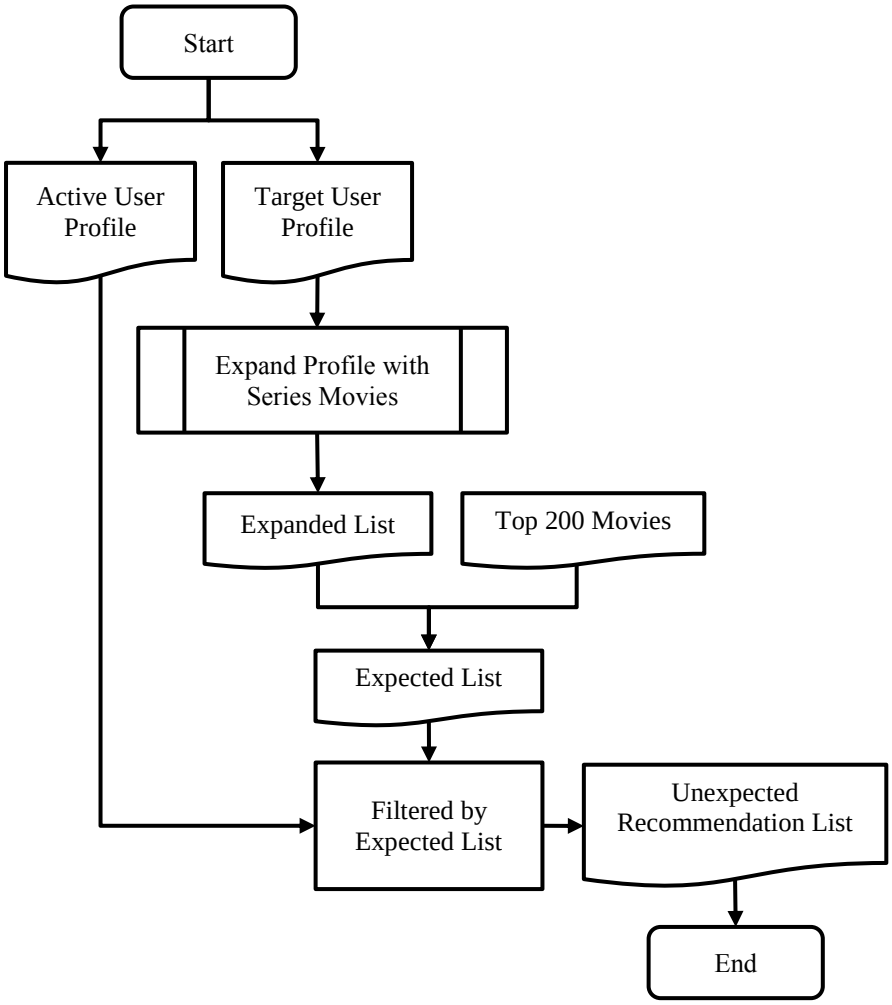


Figure 7-11 Architecture for the whole "unexpectedness" block

Obtaining expected movie list

Following the approaches proposed by Murakami et al. (2007) and Ge et al. (2010), the unexpectedness (UNEXP) list for a recommendation set can be defined as:

$$UNEXP = RS \setminus PM \quad (12)$$

where PM is a set of recommendations generated by a primitive prediction model, such as predicting items based on users' favourite categories. RS denotes the recommendations generated by a recommender system. When an element of RS does not belong to PM, this element is considered to be unexpected.

However, as argued by Adamopoulos and Tuzhilin (2015), such measures failed to fully capture the unexpectedness of recommendations as PM only contains the most popular items and does not consider user's expectations. Following the suggestions from Adamopoulos and Tuzhilin, I first define four notions during this implementation process:

$$\textit{Expected Movie List (EML): } EML = PM \cup E_u \tag{13}$$

$$\textit{Ratio of EML } (R_{EML})\textit{: } R_{EML} = \frac{N_{EML}}{N_{RS}} \tag{14}$$

$$\textit{Unexpected Movie List (UEML): } UEML = RS \backslash EMU \tag{15}$$

$$\textit{Ratio of UEML } (R_{UEML})\textit{: } R_{UEML} = \frac{N_{UEML}}{N_{RS}} \tag{16}$$

Where N_{EML} represents length of expected movie list, N_{UEML} represents the length of unexpected movie list, N_{RS} represents the length of the whole recommendation set, E_u represents user's expected dataset, for example, there is one movie of "The Lord of the Rings" in a user's profile, and then the rest two films of Lord of Ring in the same movie serious are considered as expected by the user.

Based on the above discussion, during the implementation process, two steps of work were conducted to obtain the "expected movie list" (See Figure 7-12).

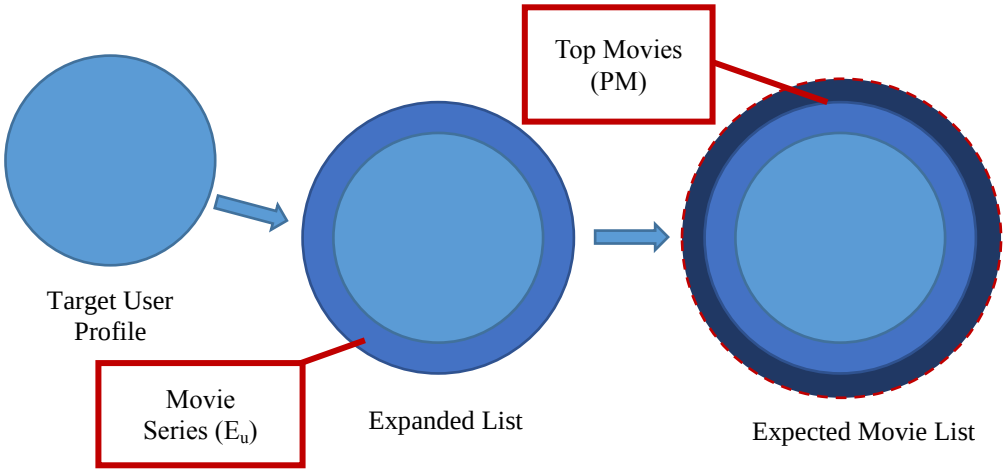


Figure 7-12 Two steps to obtain "expected movie list"

First, each target's movie profile is expanded through movie series. A premise of conducting this work is to obtain a movie collection which contains all different series of movies. This part of work is finished by following the introduction of Cormen (2009), and the way to calculate the similarity between the names of two movies is schemed as:

$$Sim(A,B) = \frac{size(LCS)}{size(longer(A,B))} \quad (17)$$

where LCS refers to the Longest Common Subsequence between string A and string B. Figure 7-13 shows the pseudo-code to perform this part of task.

Algorithm: Series movies collection

Input: The whole movie set M **;**

Output: The whole movies collection which are merged by the series S **;**

S_m : The set contains all the series movies with respect to the movie m ;

$Sim(A,B)$: A function to calculate the name similarity between the movies A and B;

for *each movie* m *in movie set* M **do**

  for *each rest movie* m' *in movie set* M **do**

    if $Sim(m,m') \geq 80\%$ **then**

    Add the movie m' into the series collection for the movie S_m.

    end

  end

  if S_m *is not empty* **then**

    Remove all the movies in S_m from M.

  end

end

Merge all the movie collections S_m into S ;

Return The whole movies collection which are merged by the series S ;

Figure 7-13 Pseudo-code for the collection of movie series

When the collection for all movie series is obtained, the target user's profile can be expanded. Figure 7-14 introduced the workflow of this part. For each movie in the target user's profile, if it is within any movie series, then the movies in the whole series are added into the expanded list; or add this movie alone to the expanded list.

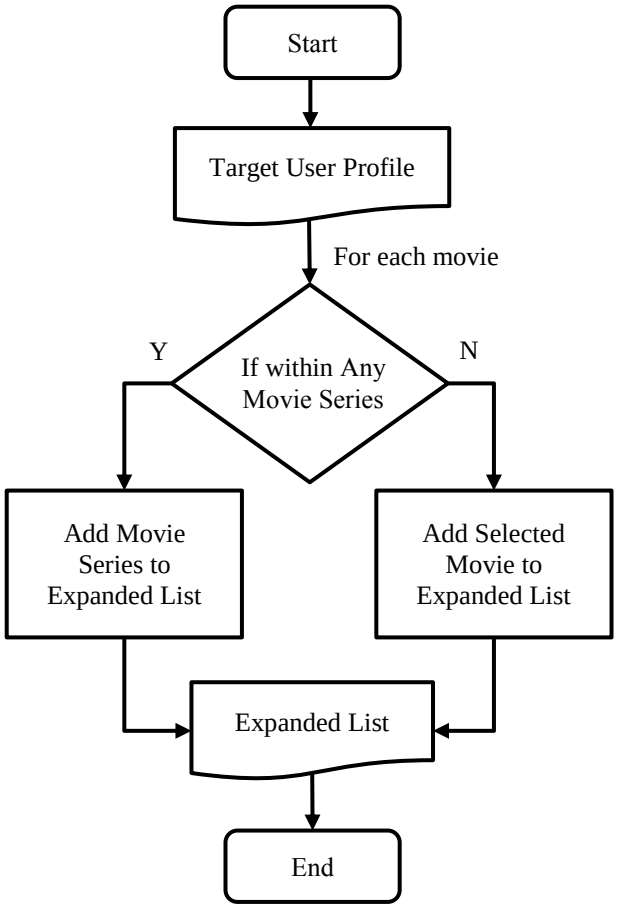


Figure 7-14 Workflow for expanding target user's profile with movie series

The next step is to add the most popular movies into the expanded list (PM). The top two hundred popular movies in the whole movie dataset are selected through the algorithm introduced in Figure 7-15. The "expected movie list" is obtained as the result of the two steps.

Algorithm: Top Two Hundred Movies Selections

Input: Rating logs L **;**

Output: Top two hundred movie set M_{top} **;**

M : The movie set which contains all the movies;

C : The counter set which counters rating logs of the movies;

for *each movie* m *in* M **do**

  for *each log* l *in* L **do**

    if $m \in l$ **then**

      Increase the specific counter c in the counter set C for movie m;

    end

  end

end

Sort the movies M into descending order based on the counter set C;

Set the first two hundred movies in M as M_{top};

return top two hundred movie set M_{top};

Figure 7-15 Pseudo-code for selecting top two hundred popular movies

As is shown in Figure 7-11, once the expected movie list is obtained, the active user's profile can be filtered into the list of "unexpected movies". The active user is obtained from the first "connection-making" block.

(3) Block 3: Usefulness block

Another important element of serendipity is value/usefulness. This element is also recognised by researchers in recommender system. Following Ge et al.'s (2010) suggestion, I define serendipity in this research setting as:

$$\textit{Serendipity Movie List (SML): } SML = UEML \cap USEFUL \tag{18}$$

$$\textit{Ratio of SML } (R_{SML})\textit{: } R_{SML} = \frac{N_{SML}}{N_{RS}} \tag{19}$$

where *USEFUL* denotes the set of "useful" items, N_{SML} denotes the length of the serendipity movie list.

Based on the development of the previous two blocks, the active user is now filtered with an unexpected movie list with respect to the target user. The next question here is how to find the useful items from the unexpected movie list and finally generate the serendipity recommendation list. The question is addressed by considering the predictions of all the items, and only those items with a prediction score larger than 3.0 are considered as useful items. Since the developed algorithm is collaborative filtering based and Pearson-correlation is calculated, the following formula is used to calculate the prediction of the active user *a*, on a certain item *i*:

$$P_{a,i} = \bar{r}_a + \frac{\sum_{u \in U}(r_{u,i} - \bar{r}_u) \cdot w_{a,u}}{\sum_{u \in U}|w_{a,u}|} \quad (20)$$

where $\bar{r}_a$ and $\bar{r}_u$ are the average ratings for the user *a* and user *u* on all other rated items, and $w_{a,u}$ is the weight calculated by Pearson-correlation between the user *a* and user *u*. The summations are over all the users $u \in U$ who have rated the item i.

According to the Equation 20, those neighbours of target users who rated the movies in the unexpected list need to be first picked out. This step is introduced in Figure 7-16, where the neighbours of target user who rate the movie in the unexpected list are found. It should be noted here that a threshold of 0.5 is set based on the Pearson-correlation. The reason for setting this threshold is that the larger of the Pearson-correlation, the larger similarity between the neighbour and the target user, and when doing the calculation of Equation 20, it can provide more accurate predictions to the movie. A profile of neighbour is generated as the result of this step.

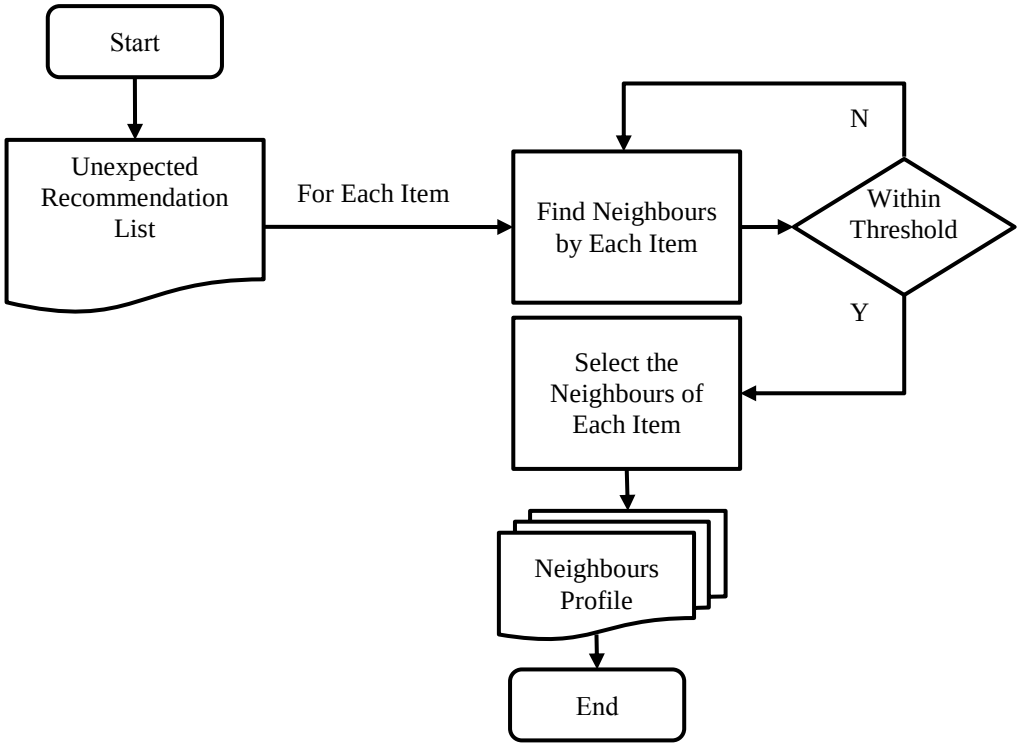


Figure 7-16 Workflow of finding neighbours before prediction

Once the neighbours are selected, the algorithm starts the prediction of each movie in the unexpected list based on Equation 20, and only those predictions larger than 3.0 (>3.0) can be expanded into serendipitous recommendation list. In the end, a serendipity recommendation list is generated on basis of the active user's profile. See Figure 7-17 as the architecture of the whole "usefulness block", as well as the pseudo-code of this block in Figure 7-18.

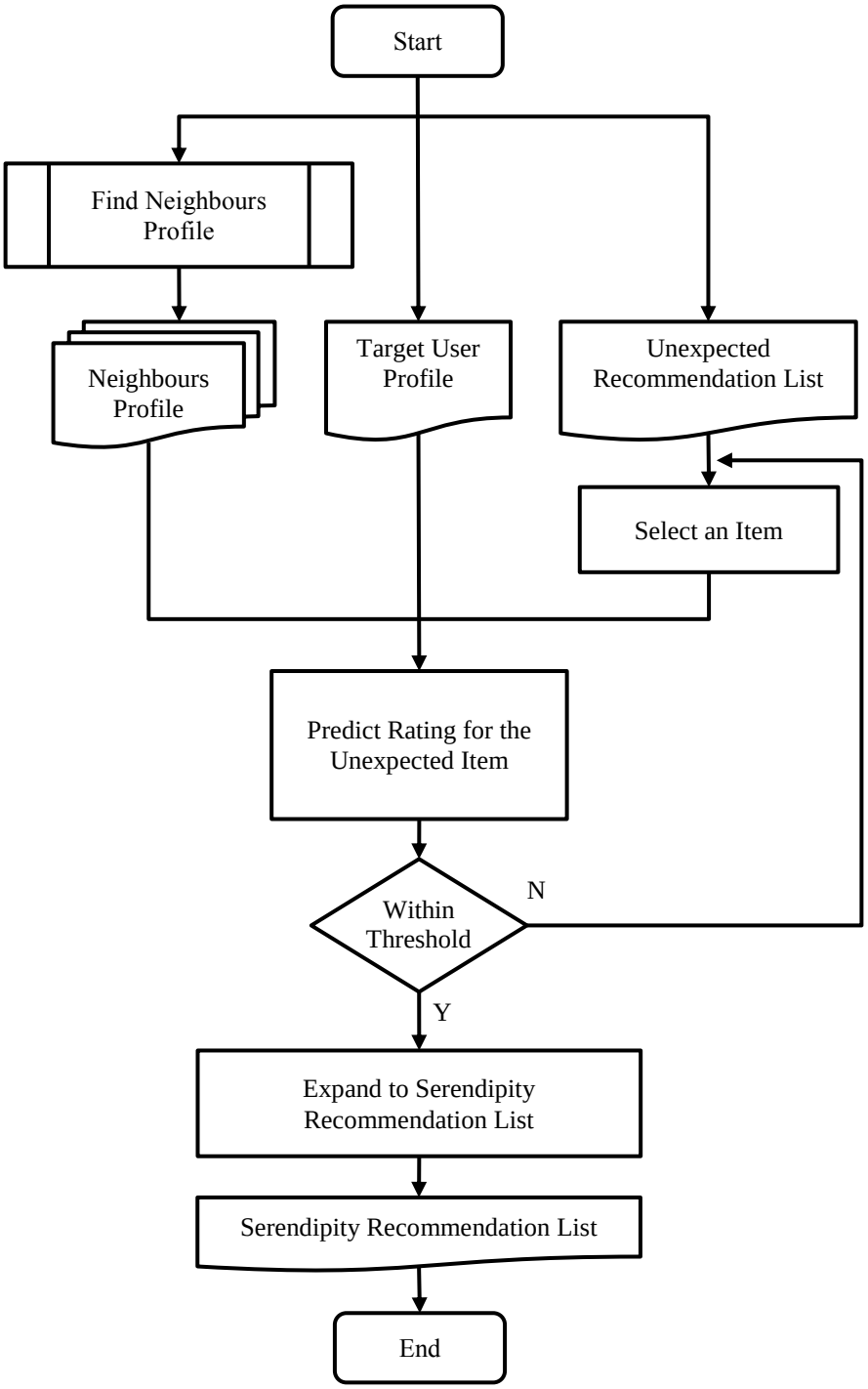


Figure 7-17 Architecture of the whole “usefulness block”

Algorithm: Usefulness filter

Input: An unexpected movie list $M_{unexpectedness}$**;**

Output: The serendipity list $SerenL$**;**

$P_{m,neighbours}$: Neighbours' profile set based on the movie m;

P_{target} : The target user's profile;

R_p : The predicted rating;

for *each movie* m *in* $M_{unexpectedness}$ **do**

Collect the neighbours' profiles $P_{m,neighbours}$ based on the movie m;

Predict the rating R_p on the movie m based on $P_{m,neighbours}$ and P_{target};

if $R_p > 3.0$ **then**

Add this movie m into the serendipity list $SerenL$;

end

end

return the serendipity list $SerenL$;

Figure 7-18 Pseudo-code of the whole "usefulness block"

Figure 7-19 depicts the topological relationship for the developed algorithm, it mainly involves three stages to generate a serendipity movie recommendation list to a target User A from User B's profile:

Stage 1: For a target User A, the developed Block 2 can help to expand user's movie list based on Equation 13, and therefore obtain an expected movie list for User A;

Stage 2: based on the developed Block 1, User B is selected the active user to

generate movie recommendation list to User A. Based on Equation 14, the User B's movie list can be separated into two parts: the overlapped part represents the expected movie list by User A, and the rest part is the unexpected movie list by User A.

Stage 3: based on the developed Block 3, in addition with Equation 18 and Equation 20, a serendipity movie recommendation can be generated in User B's unexpected movie list. It is ranked based on the predicted scores on different movies from Equation 20, and can thus be recommended to User A.

Figure 7-20 gives the pseudo-code for the developed serendipitous algorithm:

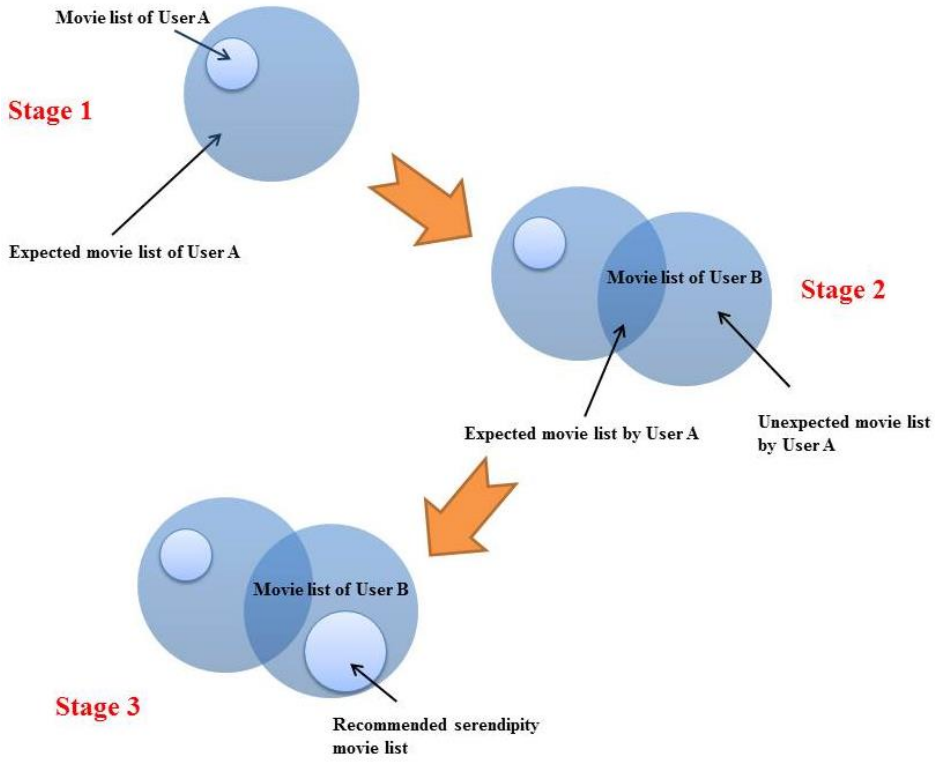


Figure 7-19 Topological relationship of the developed algorithm

Input: Target users' profile P_{target}

Output: Recommendation lists RL

U : Users except the target user for the whole data set

Th : A Pearson-correlation threshold value for filtering active users

U_{active}: Active users whose Pearson-correlations are less than or equal to the threshold

I_{active} : An item list formed from active users

$I_{popular}$: Top two hundred popular items in all items

FindSeries (i): Function that returns items in the whole item series with respect to item i (if there are)

$I_{u,unexpectedness}$: The user *u* 's unexpected item list;

$P_{i,neighbours}$: Neighbours' profile set based on the item i ;

$P_{u,target}$: The target user *u* 's profile;

R_p : The predicted rating;

$SerenL_u$: The user *u* 's serendipity list;

N_u : The user *u* 's recommendation list size;

RL_u : The user *u* 's recommendation list;

for *each target user u* **do**

Filter out active users U_{active} from U on basis of the threshold Th **;**

Form filtered item list I_{active} from active users U_{active};

for *each item* i_{target} *in target user's item list* I_{target} **do**

if $FindSeries(i_{target}) \in I_{active}$ **then**

Remove $FindSeries(i_{target})$ from I_{active};

end

end

for *each item* i_{active} *in filtered list* I_{active} **do**

if $i_{active} \in I_{popular}$ **then**

Remove i_{active} from I_{active};

end

end

$I_{u,unexpectedness} = I_{active}$;

for *each item* i *in* $I_{u,unexpectedness}$ **do**

Collect the neighbours' profiles $P_{i,neighbours}$ based on the item i;

Predict the rating R_p on the item i based on $P_{i,neighbours}$ and $P_{u,target}$;

if $R_p > 3.0$ **then**

Add this item i into the serendipity list $SerenL_u$;

end

end

Generate the recommendation list RL_u by sorting $SerenL_u$ from a high to low prediction;

Recommend RL_u to user *u* with top N_u items;

end

Figure 7-20 Pseudo-code for implementing the serendipitous algorithm

7.4 Evaluation of the Algorithm

Two experiments were designed to validate the proposed serendipitous algorithm and the developed recommender system:

- An in vitro experiment (de Gemmis et al., 2015) on a benchmark dataset, in which unexpectedness and serendipity are measured through Equation 16 and Equation 19.
- A case study with real users aiming at assessing whether the developed algorithm can effectively provide serendipitous encounters.

7.4.1 In vitro experiment

This section evaluates serendipity of recommendations on a "real-world" dataset, and the result is compared with another two most popular used collaborative filtering methods, k-nearest neighbour user-based collaborative filtering (KNN-CF) (Resnick, Iacovou, Suchak, Bergstrom, and Riedl, 1994) and item-based collaborative-based filtering (ICF) (Sarwar, Karypis, Konstan, and Riedl, 2001). These two filtering methods are widely used in commercial such as Tapestry, Amazon, etc. (Lee and Lee, 2015).

Due to the popular use of KNN-CF and ICF, during this study, Mahout was used to run these two algorithms directly. Mahout is an open source project of the Apache Software Foundation to provide developer with free program libraries. These libraries are mainly implemented with algorithms of three areas, collaborative filtering, clustering and classification. The library of collaborative filtering in Mahout is used to run KNN-CF and ICF. The pseudo-codes for the two filtering approaches are introduced in Figure 7-21 and Figure 7-22 respectively.

Algorithm: k-nearest neighbour collaborative filtering

Input: Users' profiles U

Output: User-based recommendation list RL

P_u : A user *u* 's profile;

$L_{u,sim}$: A list which contains similarity value between user *u* and others;

$L_{u,neighbour}$: A list which contains all the neighbours of the user *u*;

$L_{predicted}$: The item list with items' predicted ratings;

for *each user u in* U **do**

Collect the user *u* 's profile P_u ;

Calculate a list $L_{u,sim}$ with similarity values between user *u* and all other users;

Form user *u* 's neighbour list $L_{u,neighbour}$ by sorting $L_{u,sim}$ from a high to low order;

Generate user *u* 's predicted rating list $L_{predicted}$ in view of user *u* 's neighbour list $L_{u,neighbour}$;

Sorting the predicted rating list $L_{predicted}$ from a high to low order;

Form the recommendation list RL by the sorted predicted rating list $L_{predicted}$;

end

Figure 7-21 Pseudo-code for the KNNCF

Algorithm: Item-based collaborative filtering

Input: Users' profiles U

Output: Item-based recommendation list RL

$M_{u,i}$: The rating matrix whose rows represents users and columns represents items;

$S_{i,j}$: The list whose elements represent the similarity between item *i* and item *j*;

$M_{similarity}$: The similarity matrix whose rows and columns represent items;

$L_{predicted}$: The item list with items' predicted ratings;

for *each user u in* U **do**

Form the user-item matrix $M_{u,i}$ on basis of user *u* 's profile;

Calculate the similarity values $S_{i,j}$ between user-item matrix $M_{u,i}$;

Form the similarity matrix $M_{similarity}$ based on the similarity values $S_{i,j}$;

Generate the predicted rating list $L_{predicted}$ based on the the similarity matrix $M_{similarity}$;

Sorting the predicted rating list $L_{predicted}$ from a high to low order;

Form the recommendation list RL by the sorted predicted rating list $L_{predicted}$;

end

Figure 7-22 pseudo-code for the ICF

Measure of accuracy

Mean Absolute Error (MAE) is the mostly used metric in collaborative filtering to evaluate the accuracy of rating and item prediction (Herlocker, et al., 2004), and it is also applied in this experiment:

$$MAE = \frac{\sum_{\{i,j\}} |p_{i,j} - r_{i,j}|}{n} \quad (21)$$

where *n* is the total number of ratings over all users, $p_{i,j}$ is the predicted rating for user *i* on item *j*, and $r_{i,j}$ is the actual rating. The lower of the MAE, the better of the prediction.

7.4.1.1 Results

The serendipitous algorithm is compared with KNN-CF and ICF. In particular, three different thresholds of 0.3, 0.5 and 0.7 are compared for the developed

serendipitous algorithm. Based on Equation 16, the unexpectedness of the generated recommendation for each algorithm is caculated, and serendipity is calculated through Equation 19 while Equation 21 is used to calculate MAE of these algorithms.

Unexpectedness

Figure 7-23 depicts the unexpectedness of serendipitous algorithm with three different thresholds of 0.3, 0.5 and 0.7. According to the figure, the unexpectedness of recommendation decreases with the increasing recommender list size. It can be found that the curves with the threshold of 0.5 and 0.7 are almost overlapped, which implies there is no large difference of unexpectedness between the two thresholds. In terms of the threshold of 0.3, the unexpectedness of recommendation is similar to the other two thresholds when the recommend list size is within 400; however, the unexpectedness decreases faster than the other two when the recommending list size continues.

Figure 7-24 compares the serendipitous algorithm with the other two algorithms in terms of unexpectedness. It can be found that regardless of the slight differences among the three different thresholds for the serendipitous algorithm, the unexpectedness is obviously larger than the two most widely employed collaborative filtering algorithms and the user-based collaborative filtering leads to the least unexpectedness among the three different algorithms.

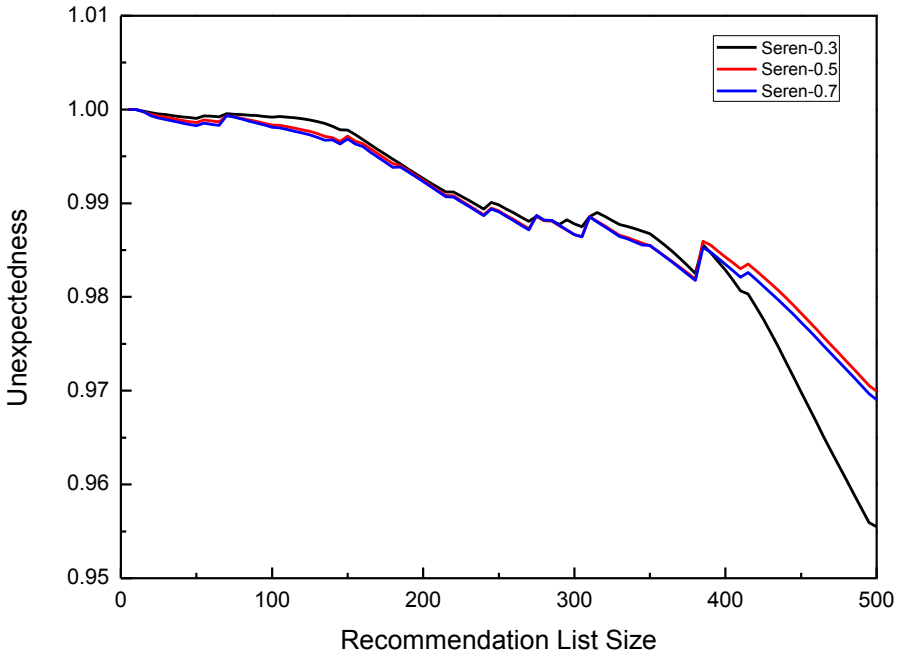


Figure 7-23 Unexpectedness of the three different thresholds for serendipitous algorithm

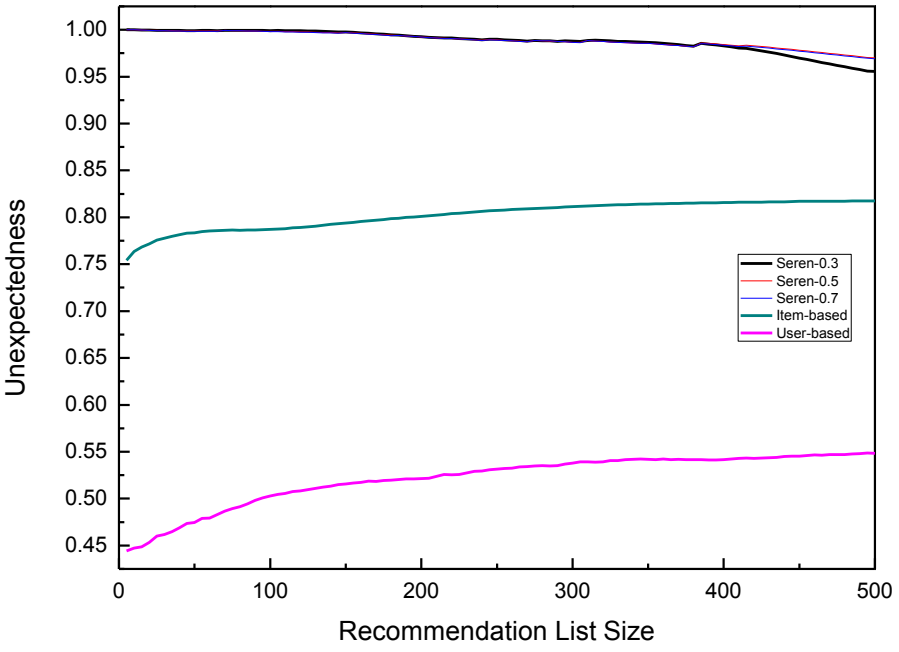


Figure 7-24 Unexpectedness of different algorithms

Serendipity

Figure 7-25 shows the calculated serendipity of the different algorithms based on Equation 19. It can be found that item-based filtering achieves the least serendipity, user-based collaborative filtering is better than item-based filtering, however, these two conventional algorithms performed much less than the proposed serendipitous algorithm. A further look into the serendipitous algorithm among the three different thresholds, it can be found that they all decrease with the recommend list size. However, comparing to the value of unexpectedness, it can be found that this time the value of serendipity deceases faster. This is mainly because of the condition of "useful", which requires the predicted rating of the movies must larger than 3.0 (Adamopoulos and Tuzhilin, 2011). Almost no differences exist among the three thresholds for serendipitous algorithm when the recommend list size is less than 100; however, a further increasing of the recommend list size, the two thresholds of 0.5 and 0.7 performs better than the threshold of 0.3. The value for serendipity of the threshold 0.5 and 0.7 are similar to the value of unexpectedness, the two curves are almost overlapped.

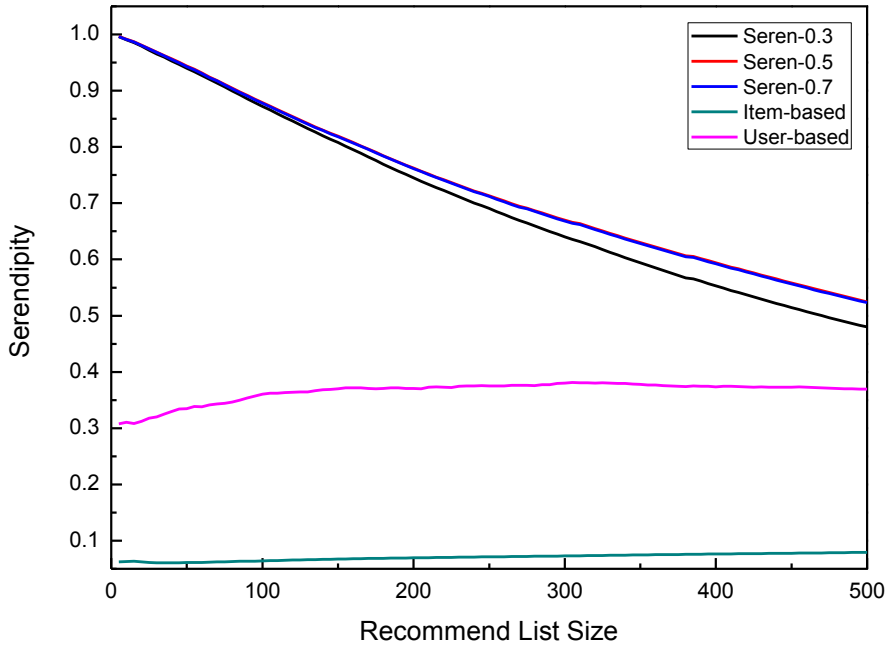


Figure 7-25 Serendipity of the different algorithms

Accuracy

The accuracy of perdition in terms of the three different algorithms is measured through MAE, and is shown in Figure 7-26. According to the figure, the developed algorithms with all the three thresholds (0.3, 0.5 and 0.7) perform better than the KNN-CF and ICF. The KNN-CF reaches a largest MAE with 0.502, while the serendipity algorithm with threshold 0.3 reaches the smallest MAE of 0.478, and the threshold of 0.5 and 0.7 are with the similar MAE of 0.480, while ICF is with a MAE of 0.486. Overall, there is no significant MAE differences exist among the three different thresholds for serendipitous algorithm.

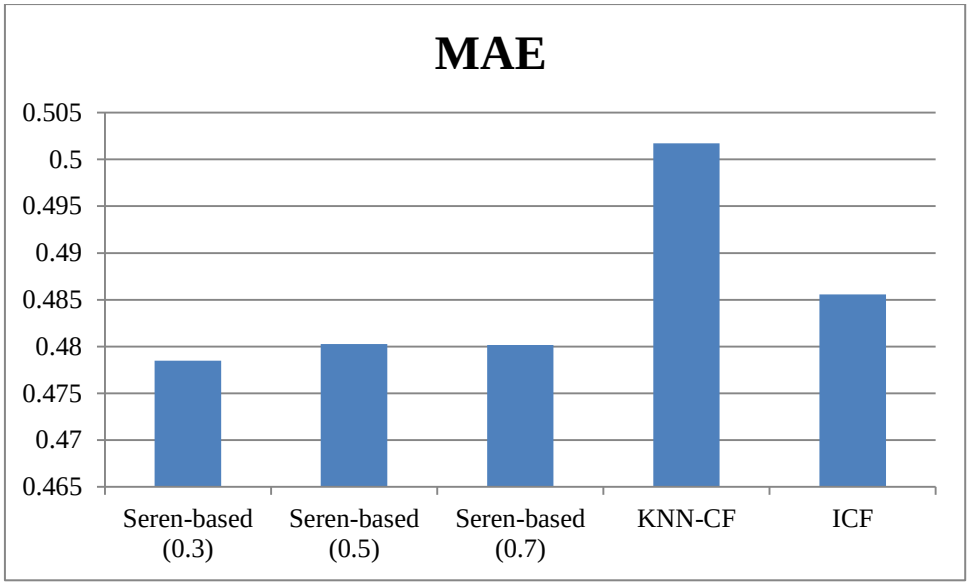


Figure 7-26 MAE of the different algorithms

7.4.2 User evaluation study

The aim of the real user study is to assess the acceptance of the recommendations produced by the developed serendipitous algorithm on real users.

7.4.2.1 Participants and dataset

12 Chinese movie scholars were invited to the study. The reason for inviting movie scholars is that a majority of the movies in the Movielens dataset were English movies, and compare to the Chinese public, they are more familiar with English movies, either from the number of movies that have been watched, or from the acceptance of these movies. Each participant has at least three years' research experience in movies. The demographic of the participants is listed in Table 7-1.

Considering the limited number of movies in the dataset introduced in Section 7.3, a larger dataset was selected to implement the developed algorithm, which is a MOVIELENS-20M dataset, which involves 138,493 users with about 20,000,263

ratings across 27,278 movies. This dataset was generated by the Movielens research group on March 31, 2015, and updated latest data to October 17, 2016. In addition, considering the performance of the algorithm as introduced in Section7.4.1.4, the threshold 0.5 was selected to run this study.

Table 7-1 Demographics of participants

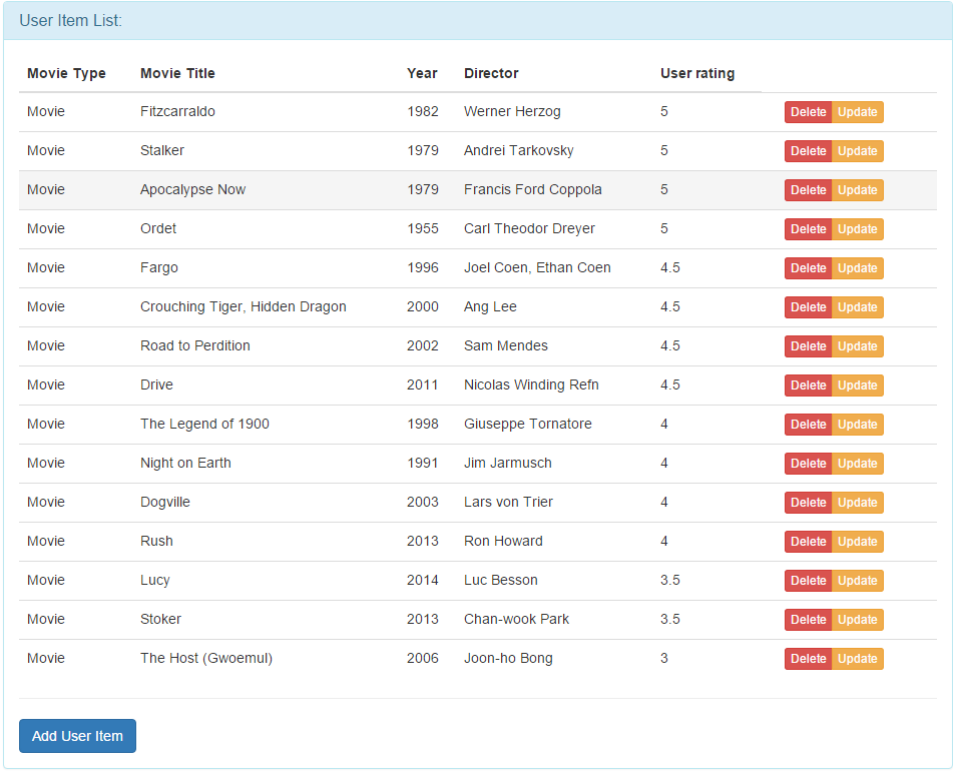
Participant	**Sex**	**Research Year on Movies**
P1	Male	10
P2	Male	4
P3	Male	10
P4	Male	4
P5	Male	3
P6	Male	9
P7	Female	5
P8	Female	5
P9	Female	12
P10	Female	6
P11	Female	3
P12	Female	6

7.4.2.2 Study platform

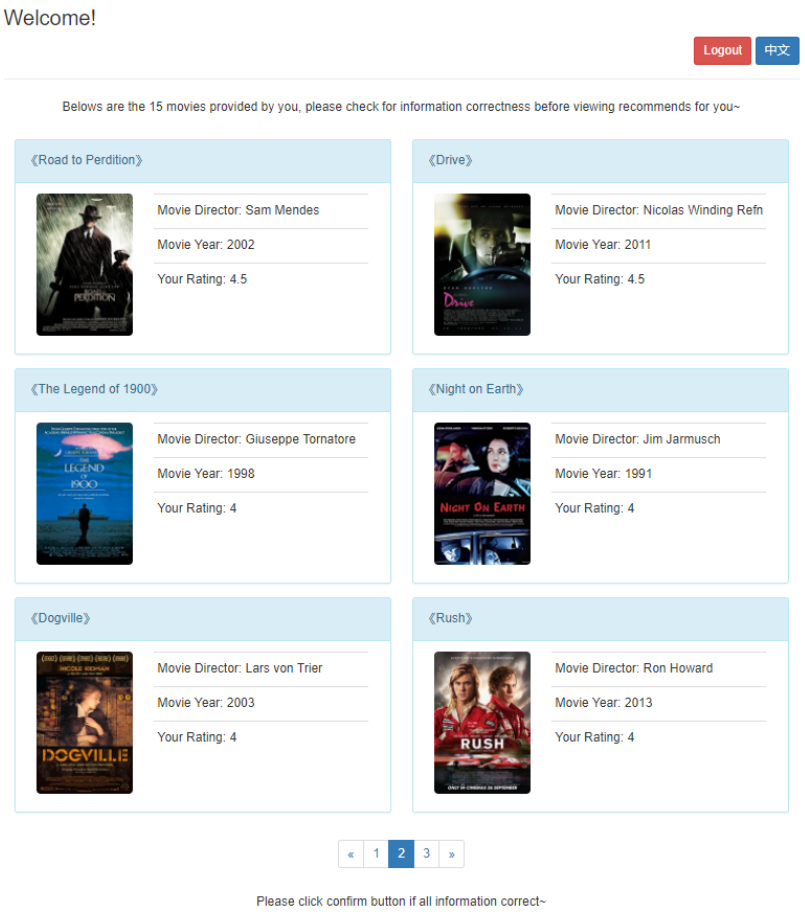
The proposed algorithm is prevalent for users with existing profiles in the movie dataset. To conduct a real world user study, a confronted challenge is to generate a profile for a new user of the dataset. Therefore a Wizard of Oz approach was applied by constructing a website to perform the user study.

The website is constructed based on a Client-Server structure. It has two modules: one is for the user, which contains functions to get access with user's profiles and

collect user's interactions on the website (e.g. inputs, logs); the other is for the administrator with functions to manage users' profiles through CRUD (Create, Read, Update and Delete). Before the study, the module for administrator from Client side posts each user's profile to the Server. The profile includes three parts: the selected 15 movies which were provided by each participant before the study, the recommended movies based on the three different algorithms (i.e. serendipity with threshold 0.5, KNN-CF and ICF), and the recommendation reasons. All the inputs on the administrator module were displayed on the user module respectively, see Figure 7-27, Figure 7-28 and Figure 7-29 as the three different parts on both modules.

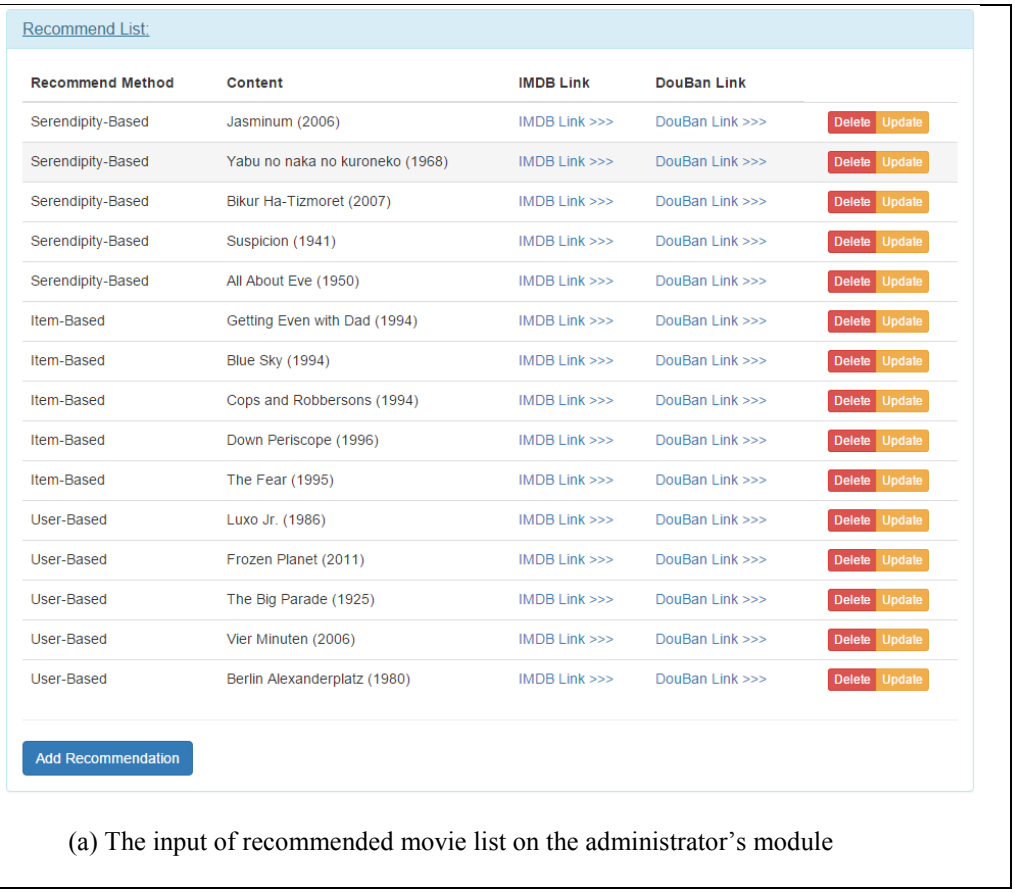


(a) The input of user's provided movie list on the administrator's module

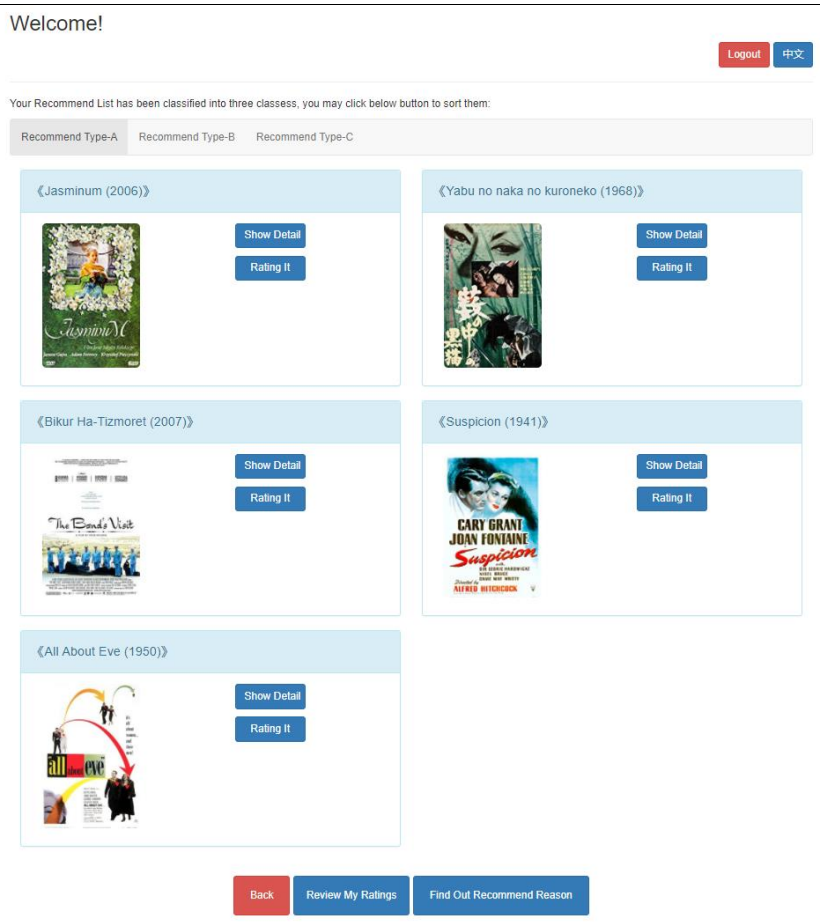


(b) The display of the input of user's provided movie list on the user's module

Figure 7-27 User's provided movie list input and its display on the two modules



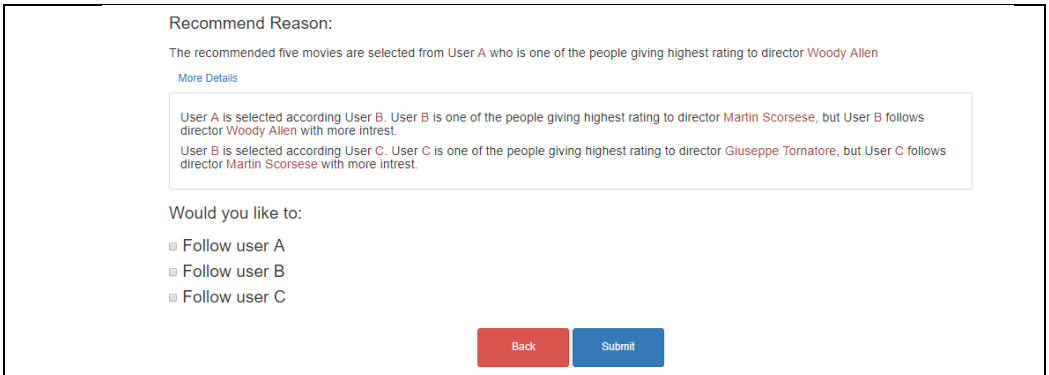
(a) The input of recommended movie list on the administrator's module



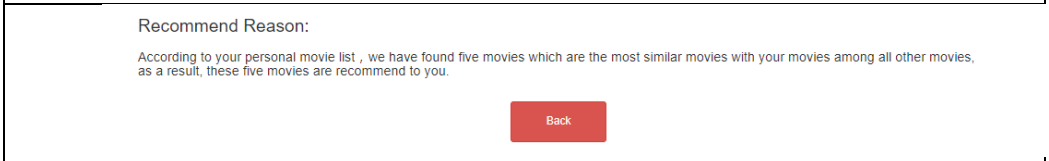
(b) The display of recommended movie list on the user's module

Figure 7-28 Recommendation movie list input and its display on the two modules

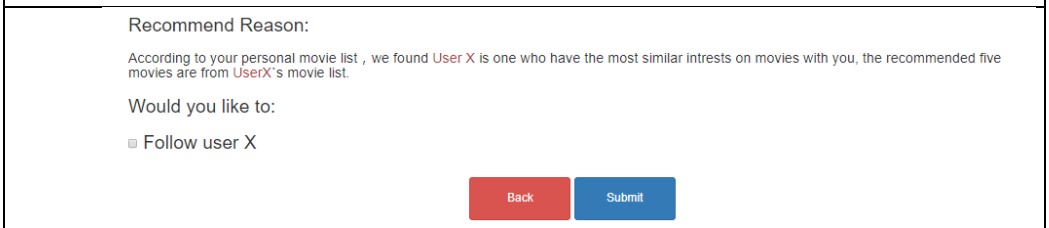
(a) The input of recommendation reason on the administrator's module



(b) The display of recommendation reason of serendipity algorithm on the user's module



(c) The display of recommendation reason of ICF on the user's module



(d) The display of recommendation reason of KNN-CF on the user's module

Figure 7-29 Recommendation reason list input and its display on the two modules

Particularly, the aim of introducing the recommendation reason in the user study is to help the target user to make connections with other information (e.g. director, most similar interest), therefore, in introducing the reason for KNN-CF (Figure 7-29(c)), a User X is set as a representation of k-nearest neighbours, which means the five movies may from one nearest neighbour or from several nearest neighbours.

The top five movies on each produced recommendation list were selected as the recommendations to participants. In particular, to avoid confusing participants of the special term of the three different recommendation descriptions, the terms of

“Recommend Type A, B and C” were used as a replacement. See Figure 7-28(b) as the example.

In addition, two functions with relating to each recommended movie were designed. The first function is a “Show Detail” button in the movie column for participants to gain detailed information of the recommended movie. Clicking on this button can link to the homepages of the recommended movie on IMDB website, where the related information of the movie is introduced in detail (e.g. director, writer, stars, storyline, user’s reviews, etc.). In addition, considering that these movie scholars are Chinese, and may have trouble in English reading, a Chinese website of Douban was also linked to each recommended movie, the website introduces the details of the movie similar to the IMDB in Chinese. Participants can choose both links, depending on their own preference (Figure 7-30).

Figure 7-30 Link the recommended movie to IMDB or Douban

The second function is a designed questionnaire which was used to measure whether the recommended movie can be serendipity in potential. The questionnaire was designed by following Ponits et al.’s (2015) study, where serendipity was evaluated through three dimensions of “unexpected”, “interesting” and “beneficial”. It is a Likert scale, ranging from “not at all” to “extremely” on each dimension, and

once a participant clicks on the “Rating it” button, the website would link to the page of questionnaire (Figure 7-31).

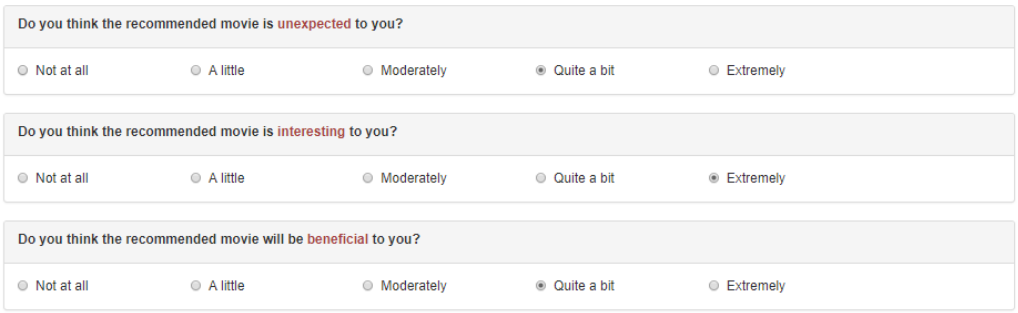


Figure 7-31 Questionnaire to measure serendipity

In particular, on the user module, a user can interact with the website by selecting which user to follow (Figure 7-29(b)) and filling into the questionnaire (Figure 7-31). During the study process, the user’s feedbacks on the two parts can be monitored by the administrator in real-time. Once the users update their feedbacks on the two parts of the study to the Server side, it can immediately obtained by the administrator on the administrator module (Figure 7-32). See Figure 7-33 as the example of user’s input of the two parts and the real-time display on the administrator module. In Figure 7-29(b), the participant chose to follow User B, it is reflected in Figure 7-33 (a), and Figure 7-33(b) shows participant’s feedback on the questionnaire.

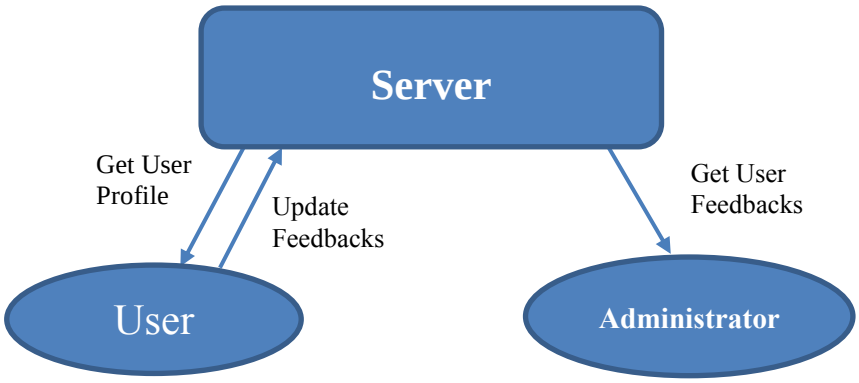
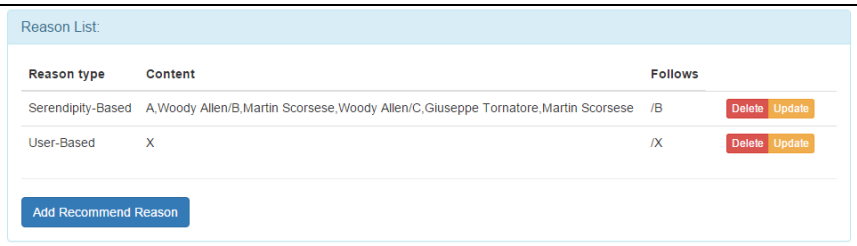


Figure 7-32 Real monitor for the administrator on user's input feedbacks



(a) The feedback on the administrator module with respect to participant's input of recommendation reason

Recommend Method	Recommend Content	Quz Content	Quz Answer
Serendipity-Based	Jasminum (2006)	Watched before	Have not watched
Serendipity-Based	Jasminum (2006)	Unexpected	Unexpected: 4 (Quite a bit)
Serendipity-Based	Jasminum (2006)	Interesting	Interesting: 5 (Extremely)
Serendipity-Based	Jasminum (2006)	Beneficial	Beneficial: 4 (Quite a bit)
Serendipity-Based	Yabu no naka no kuroneko (1968)	Watched before	Have watched
Serendipity-Based	Yabu no naka no kuroneko (1968)	Unexpected	Unexpected: 4 (Quite a bit)
Serendipity-Based	Yabu no naka no kuroneko (1968)	Interesting	Interesting: 5 (Extremely)
Serendipity-Based	Yabu no naka no kuroneko (1968)	Beneficial	Beneficial: 5 (Extremely)
Serendipity-Based	Bikur Ha-Tizmoret (2007)	Watched before	Have not watched
Serendipity-Based	Bikur Ha-Tizmoret (2007)	Unexpected	Unexpected: 5 (Extremely)
Serendipity-Based	Bikur Ha-Tizmoret (2007)	Interesting	Interesting: 5 (Extremely)
Serendipity-Based	Bikur Ha-Tizmoret (2007)	Beneficial	Beneficial: 5 (Extremely)
Serendipity-Based	Suspicion (1941)	Watched before	Have watched
Serendipity-Based	Suspicion (1941)	Unexpected	Unexpected: 3 (Moderately)
Serendipity-Based	Suspicion (1941)	Interesting	Interesting: 4 (Quite a bit)
Serendipity-Based	Suspicion (1941)	Beneficial	Beneficial: 3 (Moderately)
Serendipity-Based	All About Eve (1950)	Watched before	Have watched
Serendipity-Based	All About Eve (1950)	Unexpected	Unexpected: 3 (Moderately)
Serendipity-Based	All About Eve (1950)	Interesting	Interesting: 5 (Extremely)
Serendipity-Based	All About Eve (1950)	Beneficial	Beneficial: 3 (Moderately)

(b)The feedback on the administrator module with respect to participant's input of questionnaire

Figure 7-33 Feedback on the administrator with respect to user's input

7.4.2.3 Study procedure

Each participant was asked to provide 15 movies with their ratings three days before the study. The researcher then input all the information into the website.

Due to the participants were located in different cities across China, and online study was performed. The participant was asked to find a quiet place with internet and bring a computer with him/her when the study starts, and once the participant was ready, the researcher sent the website to the participant with a specific user name and password.

When the participant logged into the website successfully, a think-aloud method was applied during the study process. First, an online voice calling was connected between the researcher and the participant, this call was permitted from the participant before the study; the participant was then asked to speak out his/her on-going procedure, for example, when to read the movie detail, and when to fill into the questionnaire. The whole study process was under the instruction of the researcher, and participant was asked to fill into each questionnaire right after he/she finished learning about the details of the movie. This process lasted around 40 minutes for each participant to finish reading all the 15 movies under three different recommendation types and filled into the questionnaire.

A post-interview was then conducted. This interview lasted around 20 minutes and mainly focused on the reason for participant's markings of the questionnaire, as well as their preference of the three different recommendation types. The interview process was voice recorded with the permission from participants.

7.4.2.4 Data collection and analysis

Two types of data were collected from the study. The first is user's input of the quantitate data on the website, including the markings of the questionnaire and different followed users. The second type of data was the interview data.

For the first type of data, SPSS 12.0 was used to make statistical analysis and G*Power was used to conduct the power analysis. For the interview data, all the voice data was transcribed into text for data analysis. Originally, the interview was conducted by Chinese, and the data was further translated into English with the help of professional English academics.

7.4.2.5 Discussion of results

(1) Questionnaire results

An ANOVA test is performed to compare the collected marking on the questionnaire of the three different recommendation types (Table 7-2). According to the comparison, the serendipitous recommendation performs best on the dimension of unexpectedness, with a mean value of 3.28 (Std. 1.04), while user-based recommendation performed best on dimension of interesting (Mean 3.73, Std. 1.15) and item-based recommendation performed best on dimension of beneficial (Mean 3.60, Std. 1.12). However, no statistic differences exist among the three recommendation types on all the three dimensions, and the identified eta squared (η^2) are also much smaller that Cohen's guidelines of a strong effective size as 0.14 (Cohen, 1988):

- For dimension of "unexpected": F (2, 22) = 0.5, p = 0.613 > 0.05, with the effective size of η^2=0.044.

- For dimension of "interesting": F (2, 22) = 0.045, p = 0.957 > 0.05, with the effective size of η^2=0.004.

- For dimension of "beneficial": F (2, 22) = 0.157, p = 0.855 > 0.05, with the effective size of η^2=0.014.

A post-hoc power analysis is conducted based on the results from the *F*-test, and according to the feedback from G*Power, I have found that the output power (1-β error probability) for this study is weak (0.053 for the dimension of "unexpected", 0.05 for dimension of "interesting" and 0.05 for dimension of "beneficial"). Instead, a priori power analysis in G*Power showed that if I want to achieve a relatively large power of 0.8 with a strong effect size of 0.14 for this ANOVA study (Cohen, 1992), the sample size should be no less than 249.

Table 7-2 ANOVA results

	Serendipitous Recommendation		Item-based Recommendation		User-based Recommendation	
	Mean	Std	Mean	Std	Mean	Std
Unexpected	3.46	0.62	3.57	0.68	3.44	0.79
Interesting	3.16	1.09	3.61	0.80	3.58	0.73
Beneficial	3.28	0.88	3.53	0.81	3.49	0.91

Though it is a pity that no statistical differences are identified from the survey study, a good piece of news is that several participants do experience serendipity during the study. Following the measurement of serendipity by Makri and Blandford

(2012b), and Pontis et al. (2015), if the markings on all the three dimensions achieves the score of four or five (in the questionnaire, "Quite a bit" and "Extremely", see Figure 7-30), the case can be considered as an episode of serendipity or pseudo-serendipity, a term which used to describe the encounters experienced by users that have the potential of being serendipity in that users intended to do something in the future with those encounters (André et al., 2009). I also identified participant's data from this perspective (Table 7-3), and have found that more users experienced (pseudo) serendipity under serendipitous recommendation and KNN-CF, comparing to ICF; while the number of movies for the encountered (pseudo) serendipity is close to each other. The result implies participants tend to experience (pseudo) serendipity with more possibility under the serendipitous recommendation and KNN-CF.

Table 7-3 Identified serendipitous encounter

Recommendation types	**Number of users with all markings over 4**	**Number of marked movies**
Serendipity-based	8	13
ICF	5	14
KNN-CF	8	15

In addition, some participants also followed the users they were interested in after reading the recommendation reasons (Table 7-4). Comparing to KNN-CF, with only one participant preferred to follow under this situation, it can be found more participants tend to follow the introduced users under serendipitous recommendation, and there are also three participants followed both of the two recommendations. The result implies participants' preference of the recommending reasons under serendipitous recommendations.

Table 7-4 Participants choose to follow users after reading recommendation reasons

Recommendation reasons	Number of participants followed
Serendipitous-based	4
KNN-CF	1
Both	3

2. Interview results

The developed system is believed to bring more possibility for these movie scholars to experience (pseudo) serendipity. Four participants asked the researcher whether or not they can still get access with the movie information after the study, and they were quite interested to watch some of the recommended movies:

I really want to look through all the movies! I like the system as most of the [recommended] movies are new to me, and I can't wait to watch them! (Participant 8)

I love the system. Sometimes when I watched some interesting movies, I wanted to watch more, but I often don't know where and how to find the resource. The system can provide much help to me, both for my own interest or my research! (Participant 3)

I hope you can finish developing the system as soon as possible; I can't wait to use it! (P12)

All participants reported positive feedback to the function of recommendation reasons. This function helped them to understand why the system would conduct such recommendation, and provided them a chance to further explore movies from other user's files. Although there are four participants who did not follow any user in the study process, they argued this was mainly because the recommended movie did not comfort to their interest very much, and if the recommendation was interesting

enough, they would also follow other users with similar interest. The rest eight participants showed great interest to learn about similar user's profile, and three of them even expressed strong feeling to make social connections with the followed users.

However, participants also reported the recommendation reason is too simple to them currently, and this was also why more participants chose to follow users in serendipitous recommendation type, as it provided more detailed information which listed the name of the director and the relationship between the followed user and the director. However, for KNN-CF, although the system introduced the followed user shared the same interest to the participant, no further information was provided as such conclusion was obtained by calculating Pearson-correlation among different users.

7.5 Discussion and Conclusion

This chapter proposed a new method which aims to provide serendipitous recommendations to users through collaborative filtering. The algorithm is developed based on the research understanding of serendipity in information science, and the main difference between information science and recommender system with respect to the understanding on serendipity lie in the process of "connection-making". Conventionally, researchers in recommender systems usually calculate "relevance" through different algorithms in the cyberspace, such as through Pearson-correlation, to connect two users and make recommendations. However, such "relevant" connection ignores the mental process of a user's serendipitous encountering. Information researchers have demonstrated that the "making connections" can even

been considered as a kind of ability and a user must make connections through some "triggers", and these triggers are usually some kind of information that related to a user's previous knowledge/experience (Rubin, et al., 2011; McCay-Peet and Toms, 2015).

An advantage of the proposed algorithm is it pays particular attention on "triggers". As introduced in Section 7.4, the algorithm starts by finding a "familiar" standard from the user profile. Such a standard can be analysed through user profile, and in the experiment, directors are selected as the standard, other possible standards can be years, genes, actors, countries, etc. A focus on the standard can make a larger possibility for the target user to notice the relationship between the recommended user and him/herself, and thus "make connections" with the related information and finally facilitate serendipity. In addition, setting the standard provides a possibility from the system's perspective to explain reasons of recommendation. The recommender systems are conceived by users as black boxes (Swearingen and Sinha, 2001), and existing studies have showed that users are willing to learn how recommendations are generated (Herlocker, Konstan and Riedl, 2000). Lacking of recommendation reasons may even make users quit using recommender systems if the recommendations are not satisfied to them (Zaslow, 2002). This concern is addressed in the proposed serendipitous algorithm by highlighting the relationship between the active user and the target user.

While conventional collaborative filtering generates recommendations mainly depends on nearest neighbours, the proposed algorithm provides another solution to find those neighbours that is not so close to the target user by setting a threshold. The less the threshold is, the smaller possibility for conventional collaborative filtering

methods to find the neighbour. Results of the vitro experiment showed the effectiveness of the algorithm in recommending unexpected and serendipitous items to users. While the threshold of 0.5 and 0.7 outperforms 0.3, no differences existing in the two thresholds. One possible reason for such phenomenon may due to the limit size of the dataset. 90% of the Pearson-correlations are less than 0.5 among every two throughout the 2113 users in the dataset, while this number is 96% when the threshold is 0.7. In other words, only 6% of Pearson-correlations lie in the threshold from 0.5 to 0.7, and therefore no large differences can be revealed from the dataset. However, I believe the performance will be different when the dataset is large enough, and the future work will further investigate this problem on larger dataset.

No statistic differences discovered from the real user study relating to the three algorithms. This is probably due to the following two reasons: first, culture played a role and affected participants' ratings on the questionnaire. The Movielens dataset is mainly English based, and this is the reason why Chinese movie scholars are invited to the study, as comparing to other Chinese public, these scholars watched more English movies and are more acceptable of these movies. However, a reflection of the study is that a number of the movies were still new to these scholars, the related information of the movies (e.g. directors, actors) were unacknowledged to these scholars, and a simple introduction of the movie on the website was not strong enough for them to make a sufficient judgement on the three dimensions of the questionnaire. Secondly, only 12 movie scholars were invited to the user study, and the limited sample may also influence the statistical analysis of the data. The power analysis showed that it is better to have sample size no less than 249 to perform the study, however, due to the special participant status (Chinese movie scholars), it is quite

challengeable to enrol such a large sample size into the study. Future work will pay more attention on these problems by adding more movies that conforms to the culture of the participant group, and invite larger sample of participants to further compare the effectiveness of different algorithms.

An interesting discovery is that participants showed particular interest in recommendation reasons, especially the reason of serendipitous recommendation. Comparing to reasons introduced under KNN-CF, more participants chose to follow the active user profile after reading the recommendation reasons under serendipitous recommendation. Three participants even wished to make socialisation with the active users. Most participants argued that the introduction of the recommendation reason under serendipitous recommendation was more detailed than KNN-CF, and thus they were interested to follow these users. These feedbacks supported the design point of the algorithm that setting a "standard" such as directors, and the introduction of the "standard" can help participants to "make connections" with other users and finally facilitate (pseudo) serendipity.

CHAPTER 8

DISCUSSION AND CONCLUSION

8.1 Introduction

The fast developed technology in recent years have made today's people frequently encounter serendipity in digital environments (Race and Makri, 2016), and the study of serendipity thus have attracted an increasing attentions from both researchers in information science and recommender systems. Information researchers aim to discover the nature of serendipity by conducting empirical studies to investigate how users experience serendipity (e.g. Erdelez, 2004, Sun et al., 2011; McCay-Peet and Toms, 2015); while researchers in recommender systems have found that merely recommending accurate information to today's users can no longer satisfy their needs, and thus different algorisms are designed with the objective to make the recommendations more serendipitous to users (Abbassi, Amer-Yahia, Lakshmanan, Vassilvitskii, & Yu, 2009; Bhandari, Sugiyama, Datta, & Jindal, 2013; Taramigkou, Bothos, Christidis, Apostolou, & Mentzas, 2013).

This thesis investigated serendipity studies in both areas and conducted a series of empirical studies to understand serendipity among Chinese scholars in the context of information research. With the outcome from information research, this thesis also developed a new information theory-based serendipitous algorithm in recommender systems, and successfully implemented the algorithm in real data setting. The

validation experiments demonstrated that both the developed algorithm and the constructed system can help to facilitate Chinese scholar's encountering of serendipity.

The main conclusions with respect to the proposed research questions in Chapter 1 were concluded as follows:

RQ1: What are existing understandings and implementations of serendipity?

This research question is addressed in Chapter 2, where I conducted a literature review in both areas of information research and recommender systems.

Definitions of serendipity are still under debate by information researchers. Although different information researchers have proposed their definition on serendipity (e.g. van Andel, 1994; Fine and Deegan, 1996; Merton, 2004; Björneborn, 2017), none of these definitions have come to a consensus. However, as Makri and Blandford (2012a) argued, "the slippery nature of the phenomenon has meant that most existing definitions fail to incorporate something important about the experience of serendipity" (p.684), and a theoretical description (rather than definition) can better help researchers to "reduce some of the ambiguity inherent in the term's definition and usage" (p. 685). Therefore, their study proposed three key elements in encountering serendipity: unexpectedness, insight and value.

Serendipity is considered as part of information behaviour. Information researcher like Erdelez (2005) and Agarwal (2015) have proposed models of information behaviour, and serendipitous information encountering is considered as an independent part of information behaviour, which is different from purposive information seeking.

A number of theoretic research models have been put forward by information researchers. I reviewed seven existing theoretical of serendipity by different

researchers, and among them five are process-orientated (Erdelez, 2004; McCay-Peet and Toms, 2010, 2015; Lawley and Tompkins, 2008; Makri and Blandford, 2012a), while the rest two is related to the essence of serendipity (Rubin et al., 2011) and the impact of context (Sun et al., 2011). These structured models greatly helped people with a deeper insight of the slippery phenomenon.

Both content-based filtering and collaborative filtering approaches have been used to study serendipity in recommender systems. Researchers in recommender systems also discovered the important role of serendipity, and both content-based filtering (e.g. Campos & de Figueiredo, 2001; Iaquinta et al. 2008; Oku & Hattori, 2011) and collaborative filtering systems (e.g. Kawamae, Sakano, & Yamada, 2009; Lee and Lee, 2013; Onuma, Tong, & Faloutsos, 2009) have been designed to make serendipitous recommendations to users. These conducted work highlighted that there are practical possibilities to support serendipity in digital environments through strategies like recommender systems.

RQ2: What research methods can be employed to understand serendipity?

The answer to this research question is mainly proposed in Chapter 3 and Chapter 4. Chapter 3 conducted a systematic literature review to investigate the employed methodologies from existing publications, and Chapter 4 conducted an expert interview to gain a deep insight into the methodological issues in studying serendipity.

A variety of research methods have been employed to study serendipity in information research. I identified 24 published journal papers relating to serendipity studies in the context of information research from 2000 to 2017, and have found that numerous research methods has been employed by these information researcher to

study serendipity, including interview (Foster & Ford, 2003; Nutefall & Ryder, 2010; Pálsdóttir, 2011; Makri & Blandford, 2012a, 2012b; Makri et al., 2014; Jiang, Liu, & Chi, 2015; McCay-Peet & Toms, 2015), survey/questionnaire (Heinström, 2007; Stewart & Basic, 2014; McCay-Peet, Toms, & Kelloway, 2015), selective blog mining (Rubin, Burkell, & Quan-Haase, 2011), observation and think-aloud (Björneborn, 2008), controlled laboratory study (Erdelez, 2004; McCay-Peet & Toms, 2011; Makri, Bhuiya, Carthy, & Owusu-Bonsu, 2015; Miwa et al., 2011), diary-based study (Sun, Sharples, & Makri, 2011; Kefalidou & Sharples, 2016), narrative & network analysis (McBirnie & Urquhart, 2011), online ethnography (Saadatmand & Kumpulainen, 2013) and Wizard of Oz (Pontis et al., 2016). The following conducted expert interview discussed the pros and cons of employing these different methods into serendipity study.

Three methodological challenges are confronted in current serendipity study. During the expert interview, the experts discussed three main methodological challenges in studying serendipity. First, current methods lack an investigation into the instance when serendipity occurs. Second, there is a lack of consensus on the definition of serendipity, which makes it difficult for serendipity researchers to identify serendipity cases from participant's reporting, and also poses challenge to design and prepare a study. Third, there are individual differences in experiencing serendipity. For example, the ability to memory is different to each individual, and it may impact participant's description of the serendipity case during an interview study. In addition, the frequency to encounter serendipity can range the participants from super-encounterers to non-encounterers (Erdelze, 1997).

RQ3: What is the role of context played in Chinese scholars' experiencing of serendipity?

Chapter 5 and Chapter 6 conducted two user studies to address this question. In Chapter 5, a diary-based study is conducted to understand Chinese scholar's serendipitous encountering and Chapter 6 further investigated the role of emotion in experiencing serendipity.

There are three main processes for encountering serendipity. The result from the diary-based study demonstrated that Chinese scholars' serendipitous encountering mainly involves three processes: first encounter *unexpectedness* by any unforeseen means and/or unexpected content of information, then with a *connection-making* process between the encountered unexpectedness and precipitating conditions (i.e. visceral need, conscious need and previous knowledge/experience), finally recognise the *value* (i.e. substantial value and emotional value) of the encountering. Once the participant experiences the three processes, serendipity occurs.

External context, social context and internal context plays different roles in the processes of serendipitous encountering. A new context-based model of serendipity is constructed based on the diary-based study (Figure 8-1), which highlights the role of context in experiencing serendipity. The external contextual factors involve time, location and personal status; while the internal contextual factors involve precipitating conditions, sagacity/perceptiveness and emotion. The process of encountering unexpectedness is dependent on the external context and social context. The connection-making process mainly relies on the information encounterer's sagacity/perceptiveness, but sometimes can be facilitated by external context or social

context. The process of recognising value is impacted by emotions, and emotion can also influence a participant's sagacity/perceptiveness.

Emotion plays a role in encountering serendipity. In Figure 8-1, emotion is highlighted as an important internal contextual factor that can impact the process of experiencing serendipity. In Chapter 6, the role of emotion is further explored in a controlled laboratory study. By setting a sketch game as the research setting, which embedded the developed new information theory-based serendipitous algorithm, in addition with the help from the Eye-tracker and EDA sensors, the study has found that a positive emotion (success in drawing) and/or experiencing an emotional stimulus (with SCRs) were conducive to participants' scoring high marks on the evaluated dimensions of serendipity, which implies there is larger possibility for participants to encounter serendipity under such conditions.

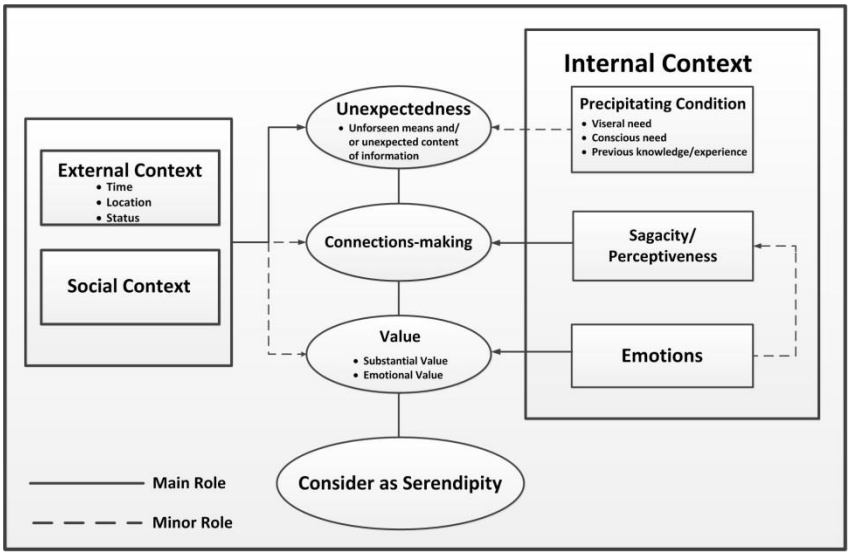


Figure 8-1 An updated context model of serendipity

RQ4: How to design recommendation technology to facilitate Chinese scholar's serendipitous encountering?

This research question is addressed in Chapter 6 and Chapter 7.

Combine information theory on serendipity to the design of algorithm in recommender systems. Based on the identified information theory from the empirical study, a new information-theory algorithm in recommender systems is developed. This algorithm combines the three key processes of encountering serendipity from the findings of the information research (i.e. encountering unexpectedness, connection-making and recognising the value). The three elements were also developed as three different blocks during the implementation of the algorithm into real data setting (i.e. "connection-making block", "unexpectedness block" and "usefulness block"). This part is introduced in detail in Chapter 7.

A validation of the algorithm on different platforms. The effectiveness of the developed algorithm is validated in both Chapters: first, in Chapter 6, the algorithm is embedded into a sketch game through a Wizard of Oz approach. Three conditions were designed to recommend different picture information to the user, and a comparison is made between the serendipitous conditions (condition 1 and condition 2) and the controlled condition (condition 3). The results from the study showed that comparing to condition 3, which is a conventional introduction of "picture-and-information", participant's scores on the four dimensions to measure serendipity (i.e. unexpectedness, interesting, relevant and beneficial) is larger under serendipitous conditions, and it implies the effectiveness of embedding the algorithm in a game setting to facilitate participants to encounter serendipity.

In Chapter 7, the algorithm is implemented in Movielens dataset, and is further evaluated through a vitro experiment in the cyberspace and a case study on real users. The vitro experiment evaluated "unexpectedness" and "serendipity" with two widely

used collaborative filtering approaches (i.e. user-based collaborative filtering and item-based collaborative filtering), and the result shows that the developed algorithm can provide recommendations with high unexpectedness and high serendipity to participants. In the case study, the algorithm is evaluated by 12 Chinese movie scholars on a self-constructed website, and participant's feedback from the study illustrates the developed serendipitous algorithm can help to facilitate serendipitous encountering.

8.2 Overview of Significant Contributions

8.2.1 Contribution 1: An initial probe of culture in studying serendipity

One of the contributions throughout the whole thesis is that I gained an understanding on how Chinese scholars experience serendipity. While existing research findings of serendipity in information research literatures are mainly dependent on Westerners, there is a lack of identification whether these findings are also pervasive to other cultural groups such as China. This thesis contributes to this research gap by carrying out a serious of studies among Chinese scholars:

1. In Chapter 5, a new context-based model is constructed from the diary-based study (Figure 8-1). The study found that Chinese scholars also experience serendipity with three major processes: encountering unexpectedness (any unforeseen means and/or unexpected content of information), connection-making between the unexpectedness and precipitating conditions (i.e. visceral need, conscious need and previous experience/knowledge), and recognise the value. The value can either be a substantial value (e.g. find the answer to previous concerns) or an emotional value (e.g. being satisfied with previous memories). The three major processes are similar to

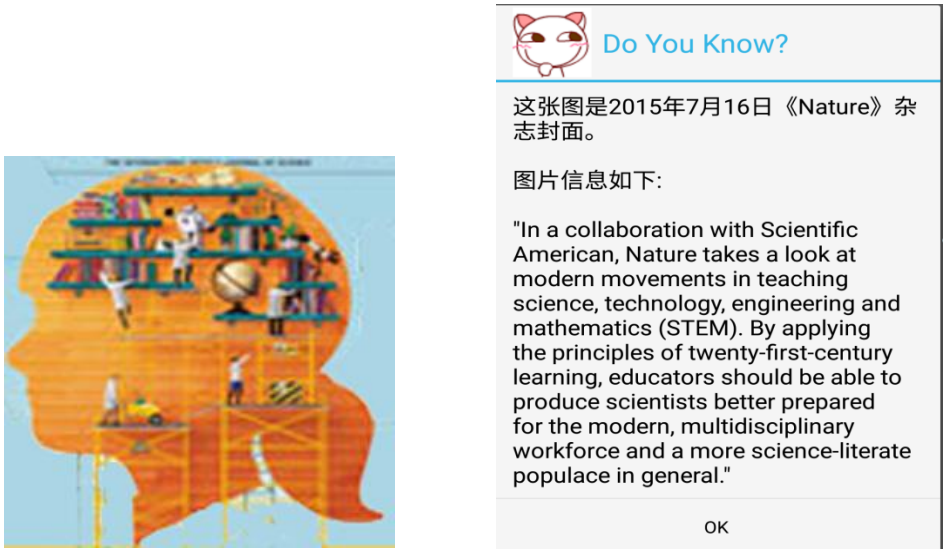
the process of serendipity discussed by other information researchers (e.g. Makri and Blandford, 2012a). This research finding provides evidence that current research understanding on the process of encountering serendipity not only is useful to Westerners, but also can be adapted to Chinese scholars.

2. There are two culture-related methodological implications with respect to studies of serendipity. The first is that there is a trade-off between the introduction of the conception for serendipity and the purpose of the study. This is reflected in the diary-based study, where participants were found with a lack of initial understanding on the term of "serendipity". Therefore the definition of serendipity in the Oxford Concise English Dictionary: "*the occurrence and development of events by chance in a happy or beneficial way*" was introduced to the participants at the beginning of the study, in addition with a collected case from a pilot study. Although during the introduction I kept highlighting to the participants that the introduced case was only one specific example, and there are other types of serendipity that can be quite different from the introduced example, such introduction of serendipity may still constrain participant's understandings of serendipity. "Serendipity" was once voted as one of the ten most difficult English words to be translated (Liu, 2013); therefore particular consideration needs to be taken by other researchers when conducting related studies on serendipity cross-culturally, especially in those countries who were initially without a basic understanding on serendipity. If the purpose of the study is to understand serendipity in a different culture group, then how can better introduce the notion of serendipity to participants? Maybe an introduction of several different types of serendipity cases from existing literatures to participants can be a possible solution.

Such diverse introduction can avoid constraining participants' understanding of serendipity to a certain situation.

The other implication is about the language used in serendipity related study, and this is mainly reflected in Chapter 6 and the case study to evaluate the developed system in Chapter 7. In Chapter 5, when conducting the diary-based study, the language of Chinese was used throughout the study, including the definition of serendipity, the design of the background picture in Wechat, and the sent reminder. There is no language problem revealed from this study. However, when performing user study 3 in Chapter 6, I designed three different conditions to embed pictures in the sketch game, with the aim to compare the developed serendipitous algorithm with the conventional "picture and information" situation. In condition 3, two pictures were selected from the "nature" website, as well as the related text introduction of the two pictures on the website (see Figure 8-2). I left the original description of picture and did not translate it into Chinese (Figure 8-2-b). However, such management resulted in participant's complains, they argued that they were preferred to read the text information in Chinese. Although all these participants were PhDs from a university where English is its official language, I still think that such non-native description of the text information may have some influence in their marking the questionnaires for serendipity evaluation. Similarly, in Chapter 7, I also conducted a case study of 12 movie scholars on a developed website, to evaluate whether or not the developed new algorithm can provide serendipitous experience to these participants. One reason to choose these movie scholars as participants is that they were much more familiar with foreign movies (e.g. English language-based movies) than the other Chinese public. However, during the study process, although I designed

both links of “Douban” and IMDB, the participants all chose to read introductions of the movie on the website of “Douban”. These two phenomena revealed throughout the whole research demonstrated that the choice of language can have influence on the study process and may probably impact participant’s experience of serendipity.



(a) Provided picture (b) Related information of the picture on the Website

Figure 8-2 Selected picture and its information from "nature" website

8.2.2 Contribution 2: Extending the understanding of serendipity

The second contribution of this thesis is that it extends existing understanding of serendipity. Instead of attempting to re-define serendipity, which has already been defined by many other researchers (e.g. van Andel 1994; McCay-Peet and Toms, 2015); this research extends current understanding on serendipity from the following three aspects:

1. An updated context-based theoretical model

As illustrated in Figure 8-1, this model is updated from the previous context model (Sun et al., 2011), and unpacked those contextual factors to different processes of serendipity. Comparing to existing models of serendipity, an advantage of this

model is the unpacked contextual factors can provide design strategies for designers to "engineer" serendipity (Bojroun, 2016). Dey et al. (2001) highlighted context is necessary "*to determine what a user is trying to accomplish. Because the user's objective is difficult to be determined directly, context cues can be used to help infer this information and to inform an application on how to best support users*". The external factors of time, location and personal status can provide design cues to facilitate serendipity, such as to make recommendations based on user's routine (time), locations (location) in leisure time (personal status). The social context implies to build social network with the group that can best support serendipity, for example, to these Chinese scholars, it is better to build socialisations with colleagues/friends. The internal contextual factors imply that designs should combine user's previous history (e.g. needs/concerns, background information) to help them to make connections with the encountered unexpectedness, and the created environment should be noticeable to users, to facilitate them to "connection-making".

2. Highlight the role of emotions during the process of serendipity

Emotion is often considered as a positive outcome that results from serendipitous encountering (e.g. Rubin et al., 2011; Sun et al., 2011; McCay-Peet and Toms, 2015), whereas current researchers lack an probe into the role of emotions played during the processes of serendipitous encountering. This thesis contributes to this research gap from two perspectives:

- As reflected in the constructed context model of serendipity (Figure 8-1), emotion is divided with two parts in a serendipitous encountering. The first is the emotional value, which is one of the recognised values by the participant; the other part is the emotion that can impact the process of evaluating the

value and an individual's sagacity/perceptiveness. This is the first time that emotion is directly put forward as an element to impact the process of serendipitous encountering.

- The second perspective is reflected by Chapter 6, where a controlled laboratory study was performed which investigated the role of emotion played in experiencing serendipity. The result from the study showed that the four dimensions (i.e. unexpectedness, relevance, interesting and beneficial) used to evaluate serendipity can be influenced by the value of emotion (positive vs. negative), and participants are more possible to experience serendipity under positive emotion. In addition, with the data collected from the EDA sensors, it can be found that the participant's occurrence of serendipity is often accompanied by SCRs.

The investigation on the role of emotions played during the process of serendipity provides new insights for designers to facilitate serendipitous encountering. "Emotional Design" which was proposed by Norman (2004) can be adapted as a design strategy, and those designs, such as "playful design" (Ferrara, 2012), "game design" (Deterding, Dixon, Khaled, & Nacke, 2011), that are able to result in a positive emotion can all be considered as possible solutions to engineer serendipity.

3. A classification of types for serendipity

Based on the identified contextual factors, especially the internal context of "precipitating conditions", this research presented a new framework to evaluate whether an encountering is serendipity or not, as is shown in Table 8-1. As discussed in Chapter 5, a key process for experiencing serendipity is connection-making between the encountered unexpectedness and precipitating conditions. Any case that

can be identified as serendipity must satisfy the three basic elements of unexpectedness, connection-making between the unexpectedness and precipitating conditions, and value. Different types of serendipity can thereby be classified under the three elements. According to my identification of the collected serendipity cases, I find a substantial value (e.g. find answer to prior concerns) is often accompanied by emotional value, and such kind of value is often resulted from those serendipity cases that relate to conscious need/concern and visceral need/concern. For those serendipity cases caused by previous experience/knowledge, it is often related with emotional value.

Table 8-1 Proposed framework to classify serendipity

		Measure of serendipity types	Yes/No
Connection-making	**Unexpectedness**	• Unforeseen means of encountering the information • Unexpected content of the encountered information	
	Precipitating conditions	• Visceral need/concern • Conscious need/concern • Previous experience/knowledge	
Value		• Substantial value • Emotional value	

8.2.3 Contribution 3: development of the new serendipitous algorithm

An important contribution of this research is the development of the new serendipitous algorithm in recommender system (See Chapter 7 for a detailed introduction of the algorithm). Though current recommender system researchers have also recognised the vital role of serendipity played in today's "information overloaded age" (Eppler & Mengis, 2004), and merely recommend accurate information to users

can no longer satisfies their demands, their study on making serendipitous recommendations are still based on an intuitive perception of the terms (Makri et al., 2014).From a user-centred design point of view, such deign for serendipity lacks empirical studies to understand how users actually experience serendipity, which is the major work studied by information researchers in recent decades.

The presented new information theory-based serendipitous algorithm in this thesis is developed based on the user study 2 (Chapter 5), hence it is user-centred and is based on the investigation of how users understand of serendipity. The algorithm is collaborative filtering-based and combines the three main elements of experiencing serendipity:

- Unexpectedness: by calculating the probability between a target user (a user to provide recommendation service) and the active user (whose user profile will be used to generate recommendations). The less of the probability, the larger unexpectedness for the target user to find the selected user.
- Connection-making: this part is implemented in the algorithm by two parts. The first is to select an attribute with a large weight in target user's profile as the start point of recommendation (e.g. a director of movies). Such a selection is with the possibility for users to make connections with the characteristics from the attribute (e.g. find a user with a relationship with the director). The other part is to show the recommendation reasons to the target user, where the attribute will be introduced to the user directly, with a demonstration of the relationship between the target user and the attribute (the real user study in Chapter 7). Such demonstration is from a design strategy that using the

recommender system to facilitate the connection-making, a process argued as an ability that can be individually different (Rubin et al., 2011; Zhou et al., 2018).

- Value: this part is also reflected from two aspects. The first is to choose the attribute with largest weight in the second step, as the larger of the weight, the larger possibility for the target user to be interested to the selected attribute, and therefore the recommendation is more possible to meet the user's interests/needs. The other part is to screen the list in the selected user file to pick typical items to generate recommendations. For example, in the movie data setting, to screen the selected user profile with predicted ratings larger than 3.

There are three major advantages of the developed algorithm. This first advantage is it jumps out the mathematical formulas to define "relevance" such as the Pearson-correlation (Resnick, Iacovou, Suchak, Bergstrom, & Riedl, 1994), cosine similarity (Sarwar, Karypis, Konstan, & Riedl, 2001), which is largely discussed by different researchers in recommender systems as an important dimension of serendipity (e.g. Yamaba, Tanoue, Takatsuka, Okazaki, & Tomita, 2013; Sun, Zhang, & Mei, 2013). Instead, this proposed algorithm is designed from a "connection-making" process in information theory on serendipity, which is more than merely a formula of "relevance". The second advantage is the developed algorithm provides a possibility to show user about the reasons behind recommendation. Recommender systems are often considered as a black boxes (Swearingen and Sinha, 2001), and an introduction of the recommender reason to users can help to improve their user experience of the systems (Herlocker, Konstan & Riedl, 2000; Zaslow, 2002). The third advantage is

that the algorithm is with theoretical support from information research, and is developed by a perception of how users experience serendipity. The successful development of the algorithm can provide implications for researchers in recommender systems, to encourage them to combine more research findings from information science when designing systems from the perspective of serendipity.

The developed algorithm is pervasive and in this PhD research, it is successfully implemented in a movie dataset. I also believe it can be adaptable to other data sets like libraries, shopping, music, etc.

8.2.4 Other contributions

Besides the three core contributions, there are also other contributions of the thesis that worth to be discussed.

(1) An exploration of research methods in serendipity study

This research contributes to new and novel approaches to observe serendipity in controlled research setting. In Chapter 6, I conducted an empirical study and successfully employed HCI equipment of Eye-tracker and EDA sensors on participants. This is a new attempt, as a literature review in Chapter 3 on current research methods showed that it is mainly those conventional research methods (e.g. interview, survey) that are employed to understand the subjective phenomenon of serendipity, and a large disadvantage identified from the expert interview is that data collected by these research methods are retrospective. Foster argues there is still a need to capture the data "at the moment" when serendipity occurs (2014). With the help of EDA sensors and Eye-tracker, participant's emotional responses during the process of serendipity were successfully captured, and I have thus found from the

study that participant's evaluation of serendipity can be impacted by emotions: participants are more possible to experience serendipity under positive emotions, and the occurrence of serendipity is often accompanied with SCRs. The collected data by the EDA sensors provide objective evidence for analysing the role of emotions played in serendipitous encountering. To the best knowledge, thus far only Miwa et al.'s (2011) work had employed Eye-tracker to their serendipity study and none of prior studies have attempted to use any physiological sensors to capture relevant data from participants. The successful attempt provides implications that through a careful design of study, it is possible for researchers to capture the "instant" moment of serendipity, and more attentions can be paid by future researchers to extend research on this direction to better understand the cryptic phenomenon of serendipity.

(2) Various research platforms

Three different research platforms are employed in the empirical studies throughout the thesis. In Chapter 5, when performing the diary-based study, "Wechat" was selected as the research platform, and this was under a deliberately comparison with the previously self-developed dairy application (Sun et al., 2011). According to Sun et al (2011), the developed dairy application need to be installed to every additional mobile device, and delivered them to each participant. They report in their paper that their participants reported that carrying such an additional device was inconvenient during the study process. When designing this diary study, I found that nowadays the fast developed technology has provided huge convenience to researchers. The social media "Wechat" covers almost all the functions of the previous diary application, and there is no need for participants to bring additional devices, they can record their serendipitous experience instantly through different

forms (i.e. text, videos, pictures and voice). In addition, these participants are all frequent users of "Wechat", no extra time is needed for them to learn about the functions of the application. Such social application can also provide convenience for researchers to send daily reminders to participants, help them to better recognise the experiment situation. Researcher can also design background pictures to help participant better record their encountering, see Table 8-2 as the advantages of the usage of "Wechat" as the platform for conducting the diary based study.

Table 8-2 Advantages of applying Wechat as research platform

Participants	**Researchers**
• No need to carry additional device • Frequent users, no need extra time to learn about the functions of the application • With the same functions to record serendipitous encountering (i.e. text, videos, pictures and voice) as previous application • Send encountered serendipity to researchers instantly • Contact researchers immediately if any problems happens during the study	• Easy to send daily reminders to better control the study • Reply to participants timely (e.g. problems, successfully receive of information) • Design background pictures to better help participants record serendipitous encountering

The second research platform employed in this research is the android-based sketch game introduced in Chapter 6. A main reason to use game is that game has the ability to result in different emotion values. A successful accomplishment of the game is often accompanied with positive emotions, while a failed accomplishment often

leads to negative emotions. According to Shute's (2011) theory of the Evident-Centred Design (ECD), a player's abilities and understandings, especially those that cannot be directly observed by researchers (e.g. critical thinking, problem solving) is reflected through the embedded tasks or situations in the design, such as the interaction processes of the game. Serendipity is exactly such a phenomenon that cannot be observed directly by the researchers; however, during the process of game-playing, participants would naturally produce sequences of actions while performing the designed tasks and hence provides possible evidences for researchers to access the encounter of serendipity. The designed task in this empirical study is based on the developed algorithm, and the result showed the effectiveness in employing the game-based experimental situation and the developed serendipitous algorithm.

The third research platform is the developed website in evaluating the constructed movie recommender systems. With the construction of the website, I compared the developed serendipitous algorithm with another two widely used user-based collaborative filtering approaches (i.e. k-nearest neighbour collaborative filtering and item-based collaborative filtering). The effectiveness of the developed serendipitous algorithm is identified during the study, which can help to facilitate participant's serendipitous encountering.

The different employed research platforms imply that exist technology can be used to support researchers to design serendipity studies. Apart from the traditional interview or survey questions, a productive serendipity study can be performed under a flexible selection of different research platforms to address different research purposes.

(3) Extending possibilities of design research setting for experiencing serendipity

It is always a challenge for information researchers to design research setting that can support participants experience serendipity. In previous studies, information researchers mainly design experiments by distributing information searching tasks to their participants (e.g. Erdelez, 2004; McCay-Peet and Toms, 2011; Makri et al., 2015), and this constrains the encountering of serendipity in the context of active information searching. Agarwal's (2015) proposed three dimensions of serendipity:

- During a passive phase (including browsing, scanning or non-purposive seeking):

1. {Accidental, incidental, serendipitous, unintentional or chance} {encountering, finding, stumbling upon, acquisition or discovery} of information.

- During an active (purposive seeking or search) phase:

2. {Accidental, incidental, serendipitous, unintentional or chance} {encountering, finding, stumbling upon, acquisition or discovery} of information.

3. {Opportunistic or active} {encountering, finding, stumbling upon, acquisition or discovery} of information.

Obviously, assigning information tasks to participant limits the occurrence of serendipity in the second and/or third dimension, as the participants were asked to perform active information seeking tasks during the experimental process, whereas the first dimension of "passive phase" is lack of investigation. However, based on the developed serendipitous algorithm and its effectiveness identified in Chapter 6 and

Chapter 7, the design of research setting for participants to experience serendipity is greatly extended: first, the context for encountering of serendipity can no longer be limited in information seeking processes. As the introduced game-based user study in Chapter 6, a game research setting is designed with the employment of the algorithm, and participants can experience serendipity also in game context, which help us to have a look into the role of emotion in serendipitous encountering. In addition, the algorithm makes researchers possible to design the first serendipity dimension proposed by Agarwal (2015), where participants experience serendipity in a "passive phase", for example, both research settings of the game study in Chapter 6 and the case study in Chapter 7 are passive to participants, and they did not conduct any active or purposive searching. I believe with the help from the developed serendipitous algorithm, future studies can design more contexts for participants to experience serendipity.

8.3 Limitations

Through a series of studies, this PhD research investigated how Chinese scholars experience serendipity and how to design environments to help Chinese scholars facilitate serendipitous encountering. However, there are also some limitations arise from the works conducted throughout the thesis.

The first is the methodological limitations. In the diary-based study introduced in Chapter 5, due to lack of an initial understanding of serendipity by participants, I introduced the notion of "serendipity" through its definition in the dictionary, in addition with a collected case from pilot study. Although I kept highlighting to the participants that the introduced case is only one type of serendipity and serendipity is

definitely not limited to this type, I still think such introduction may probably constrain the participant's recognition of the notion on serendipity. In addition, during the two-week diary study, I sent participants reminders every day. According to the feedbacks from the participants, although such reminders worked well in reminding them with an experimental situation, they may also cause pressure to some of the participants, especially those who were not frequent encounterers. In the controlled laboratory study introduced in Chapter 6, participants were asked to put on EDA sensors during the study process and under the monitor of the Eye-tracker. Such strict experiment situation can cause pressure to participants, and there was even one participant reported that he preferred to finish the study by himself without any monitor. It seems there is a trade-off between the usage of HCI equipment and the possible pressure on participants caused by the equipment. More careful considerations need to be taken to make participant with a least pressure in the experimental situation.

Second, the number of participants enrolled in all the four studies is limited. In the expert interview, I successfully enrolled some experts whose publications were identified in the systematic literature review, and they provided effective and practical suggestions on the methodology issues with respect to serendipity studies. However, it is also a pity that I failed to enrol all the experts in the identified publications in Chapter 3. I believe a more systematic overview on the methodology issues of serendipity studies can be generated if I can collect the feedbacks from all experts that with empirical study experiences in the identified publications. In the diary-based study, I enrolled sixteen Chinese scholars from a variety of research backgrounds. All these participants were enrolled from the same campus and therefore there were no

large distinctions in their research atmosphere. More identification on the Chinese scholars that are from different research conditions (e.g. different universities) are needed to have a more deep look into the serendipitous encountering among a more representative group participants of Chinese scholars. Also, in the controlled laboratory study and the case study of evaluating the developed algorithm on the website, the participants were invited to fill in the questionnaire to measure whether or not they had encountered serendipity. The limited numbers of samples in the two studies made us failed in performing strict and systematic statistical analyses, and this is probably a reason that I had not identified significant differences among different groups of data.

Third, there are limitations in the evaluation of serendipity when employing the questionnaires in Chapter 6 and Chapter 7. The two studies all used the developed questionnaire by Makri and Blandford (2012b) to evaluate the occurrence of serendipity among participants, which was also adopted in other related studies (Kelifaction & Sharples, 2016; Points et al., 2015). However, during the study process, I found that there were limitations in using the questionnaire especially on the dimension of "unexpectedness". As reported in the game-based study, there are two levels of "unexpectedness", one is the unexpected relationship between the provide picture and the text information, the other is the unexpected content of the text information. The two levels of unexpectedness made the participant biased in filling in this column of the questionnaire, and the situation was quite similar in the case study in Chapter 7, the unexpectedness existed in two levels, the relationship between the provided movie and participant's selected movie before the study, as well as the unexpected content of the recommended movie. During the study process, due to the

small number of participants, I distinguished the two levels of unexpectedness through the interview process; however, if conducting a serendipity study with large number of participants, it will be a problem for participant's rating on this dimension. A more appropriate design on the dimension of "unexpectedness" is still needed to be developed to better measure serendipity in future studies.

The last limitation is I used the dataset of Movielens to implement the developed algorithm in Chapter 7. This dataset is selected because it is a widely used dataset in recommender systems and a number of user studies have been performed on basis of this dataset (e.g. Adamopoulos & Tuzhilin, 2011; de Gemmis et al., 2015). However, during the case study, I found that most of the movies involved in the dataset are English-based movies, and there is a language concern when performing studies on the Chinese scholars. Almost all the participants chose to log onto the Chinese website of "Douban" to read the introduction of the movie. However, I have found that many of the movie introductions on "Douban" are not as detailed as their introductions on the website of IMDB. Participants may have more detailed information on the recommended movies if they can log onto IMDB during the study process, and this is a limitation of the study. Future work may need to consider implementing the algorithm in other data set which is more appropriate to user groups.

8.4 Future Work

Following from this thesis' contributions and limitations, future studies on serendipity research may be extended in several possible ways:

(1) Extend the research on culture dimension. This thesis investigates serendipitous encountering among Chinese scholars, and has an initial probe into the

role of culture. However, no strict culture comparative studies are performed in this thesis, and it can be a future direction. For example, future studies can adopt research findings from culture study (e.g. Hofstede's culture model (1980)) to design more strict and comparative studies among different cultural groups, to extend the window of serendipity study to a culture dimension.

(2) Explore more empirical studies to further validate the findings from the updated context-based model of serendipity. A major contribution of this thesis is the new updated context-based model, and different design strategies can be generated on basis of the identified contextual factors. Various empirical studies can be explored to further validate the role of these contextual factors played in encountering serendipity. For example, social context plays an important role in participants' serendipitous encountering, but what are the differences in encountering serendipity caused by different social counterparts? Which social distance (e.g. family, superior, colleague, friends, and strangers) can better facilitate participants' serendipitous encountering? In addition, while sagacity/perceptiveness is considered as internal contextual factors to impact the process of serendipity, then how to design systems enhance participant's perceptiveness? Further, since emotions can also influence the process of serendipity, the research outcome from emotional studies can also be combined to study serendipity, such as whether research models in emotion studies can be used to better help understanding the role of emotion in experiencing serendipity? What kind of emotion can contribute to the occurrence of serendipity? I believe a number of works can be extracted on basis of the developed context-based model to have a deep exploration of the nature of serendipity.

(3) A further exploration of research methodologies. In the game-based study, I made new attempts on the research methodology, the EDA sensors were used to capture participant's physiological data, and I successfully collected objective data (SCRs) during serendipitous processes in that study. Future studies can try to capture other types of physiological data by other physiological tools, such as participant's cardiovascular signals, respiration, brain activity, muscle tension, etc. (see Lie et al. (2017) for a detailed introduction). In addition, the Eye-tracker in the study is mainly used to help identify different game processes. Future studies can also try to collect the data produced by the Eye-tracker, such as eye-gaze points, pupil diameters, etc. An advantage of exploring these HCI tools into serendipity study is that they can provide solutions to collect objective data during the study process, and to help researchers with a clearer perception on participant's performances during the study.

(4) A further validation of the developed serendipitous algorithm. This thesis implemented the developed information theory-based serendipitous algorithm in a movie data setting and embedded the algorithm to a self-constructed website through a Wizard of Oz approach. Although in the case study I selected movie scholars as participants, I believe that the algorithm can be pervasive to any other user groups. Future studies will continue to conduct a series of studies to evaluate the effectiveness of the developed algorithm from three major perspectives: first, to enrol a large number of participants to the empirical study. The enrolled participants will no more limited to scholars, but will be more general user groups. With data collected from large samples, strict statistical analysis can be conducted to validate the effectiveness of the developed algorithm and system. Second, during this thesis, I compared the developed algorithm with two popular collaborative filtering approaches (i.e. user-

based collaborative filtering and item-based collaborative filtering). Future studies will also make comparisons with other proposed recommend algorithms that aim to provide serendipitous recommendations to users (e.g. Sun, Zhang, & Mei, 2013; de Gemmis, Lops, Semeraro, & Musto, 2015). A comparison with these algorithms can help us have a better perception and optimisation of the developed algorithm. Third, I will also try to plant the algorithm in other recommender settings, such as music, books, etc. to test the effectiveness of the algorithm in these different settings.

Bibliography

Abbassi, Z., Amer-Yahia, S., Lakshmanan, L. V., Vassilvitskii, S., & Yu, C. (2009). Getting recommender systems to think outside the box. In *Proceedings of the third ACM conference on Recommender systems* (pp. 285–288). ACM.

Adamopoulos, P., & Tuzhilin, A. (2011). On Unexpectedness in Recommender Systems: Or How to Expect the Unexpected. In *DiveRS@ RecSys* (pp. 11–18).

Adamopoulos, P., & Tuzhilin, A. (2015). On unexpectedness in recommender systems: Or how to better expect the unexpected. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *5*(4), 54.

Agarwal, N. K. (2015). Towards a definition of serendipity in information behaviour. *Information Research: An International Electronic Journal*, *20*(3), n3.

Als, B. S., Jensen, J. J., & Skov, M. B. (2005). Comparison of think-aloud and constructive interaction in usability testing with children. In *Proceedings of the 2005 conference on Interaction design and children* (pp. 9–16). ACM.

Andel, P. van. (1994). Anatomy of the unsought finding. serendipity: Orgin, history, domains, traditions, appearances, patterns and programmability. *The British Journal for the Philosophy of Science*, *45*(2), 631–648.

Andrę, P., & others. (2009). Designing for (un) serendipity-computing and chance. *The Biochemist*, *31*(6), 19–22.

Apted, E., & Choo, C. (1971). General purposive browsing. *Library Association Record*, *73*(12), 228–230.

AsIMov, I. (1976). Biographical Encyclopaedia of Science& Technology .New York: Avon Books, p.

Bhandari, U., Sugiyama, K., Datta, A., & Jindal, R. (2013). Serendipitous recommendation for mobile apps using item-item similarity graph. In *Asia Information Retrieval Symposium* (pp. 440–451). Springer.

Björneborn, L. (2008). Serendipity dimensions and users' information behaviour in the physical library interface. *Information Research*, *13*(4).

Björneborn, L. (2017). Three key affordances for serendipity: Toward a framework connecting environmental and personal factors in serendipitous encounters. *Journal of Documentation*, *73*(5), 1053–1081. https://doi.org/10.1108/JD-07-2016-0097

Bogers, T., Rasmussen, R. R., & Jensen, L. S. B. (2013). Measuring serendipity in the lab: The effects of priming and monitoring.

Bordino, I., Mejova, Y., & Lalmas, M. (2013). Penguins in sweaters, or serendipitous entity search on user-generated content. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management* (pp. 109–118). ACM.

Boyce, B. R., Meadow, C. T., & Kraft, D. H. (1994). *Measurement in information science*. San Diego, CA: Academic Press.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Bryman, A. (2001). The nature of qualitative research. *Social Research Methods*, 264–288.

Buchwald, N. (1999). *Thinking of information through the humanist's eyes.* Available at: http://dizzy.library.arizona.edu/users/buchwald/humanities.html (retrieved August 11, 2004).

Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., & Ito, T. A. (2000). The psychophysiology of emotion. *Handbook of emotions*, *2*, 173-191.

Campos, J., & Figueiredo, A. D. de. (2001). Searching the unsearchable: Inducing serendipitous insights.

Cantador, I., Brusilovsky, P. L., & Kuflik, T. (2011). *Second workshop on information heterogeneity and fusion in recommender systems (HetRec2011)*. ACM.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: using vision to think*. Morgan Kaufmann.

Case, D. O. (2012). *Looking for Information*. Emerald Group Pub. Ltd. ;

Case, D. O., & Given, L. M. (2016). Looking for information: a survey of research on information seeking, needs, and behavior, 4th edition. *Journal of the Medical Library Association*, *91*(2), 259.

Cobbledick, S. (1996). The information-seeking behavior of artists: Exploratory interviews. *The Library Quarterly*, *66*(4), 343–372.

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155.

Cooksey, E. B. (2004). Too important to be left to chance—Serendipity and the digital library. *Science & Technology Libraries*, *25*(1–2), 23–32.

Cormen, T. H. (2009). *Introduction to algorithms*. MIT press.

Courtright, C. (2007). Context in information behavior research. *Annual Review of Information Science and Technology*, *41*(1), 273–306.

Crow, S. R. (2011). Exploring the Experiences of Upper Elementary School Children Who Are Intrinsically Motivated to Seek Information. *School Library Media Research*, *14*, 42.

Crow, S. R. (2015). The Information-Seeking Behavior of Intrinsically Motivated Elementary School Children of a Collectivist Culture. *School Library Media Research*, *18*, 2165–1019.

Cunha, M.P. (2005), "Serendipity: why some organisations are luckier than others", FEUNL working paper series, Universidade Nova de Lisboa, Lisbon.

Cunha, M. P., Clegg, S. R., & Mendonça, S. (2010). On serendipity and organizing. *European Management Journal*, *28*(5), 319–330.

de Gemmis, M., Lops, P., Semeraro, G., & Musto, C. (2015). An investigation on the serendipity problem in recommender systems. *Information Processing & Management*, *51*(5), 695–717. https://doi.org/10.1016/j.ipm.2015.06.008

Delgadillo, R., & Lynch, B. P. (1999). Future historians: Their quest for information. *College & Research Libraries*, *60*(3), 245–259.

Dey, A. K., Abowd, G. D., & Salber, D. (2001). A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human-Computer Interaction*, *16*(2), 97–166.

El-Sheikh, M. (2007). Children's skin conductance level and reactivity: Are these measures stable over time and across tasks? *Developmental Psychobiology*, *49*(2), 180–186.

Erdelez, S. (1997). Information encountering: a conceptual framework for accidental information discovery. In *Proceedings of an international conference on Information seeking in context* (pp. 412–421). Taylor Graham Publishing.

Erdelez, S. (2004). Investigation of information encountering in the controlled research environment. *Information Processing & Management*, *40*(6), 1013–1025.

Erdelez, S. (2005). Information encountering. *Theories of Information Behavior*, 179–184.

Erdelez, S., Basic, J., & Levitov, D. D. (2011). Potential for inclusion of information encountering within information literacy models. *Information Research: An International Electronic Journal*, *16*(3), n3.

Erdelez, S., Heinström, J., Makri, S., Björneborn, L., Beheshti, J., Toms, E., & Agarwal, N. K. (2016). Research perspectives on serendipity and information encountering. *Proceedings of the Association for Information Science and Technology*, *53*(1), 1–5.

Fine, G. A., & Deegan, J. G. (1996). Three principles of serendip: insight, chance, and discovery in qualitative research. *International Journal of Qualitative Studies in Education*, *9*(4), 434–447.

Foster, A. (2004). A nonlinear model of information-seeking behavior. *Journal of the Association for Information Science and Technology*, *55*(3), 228–237.

Foster, A., & Ellis, D. (2014). Serendipity and its study. *Journal of Documentation*, *70*(6), 1015–1038.

Foster, A., & Ford, N. (2003). Serendipity and information seeking: an empirical study. *Journal of Documentation*, *59*(3), 321–340.

Fredrickson, B. L. (1998). What good are positive emotions?. *Review of general psychology*, *2*(3), 300.

Fredrickson, B. L., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition & Emotion*, *19*(3), 313–332.

Fröberg, J. E. (1977). Twenty-four-hour patterns in human performance, subjective and physiological variables and differences between morning and evening active subjects. *Biological Psychology*, *5*(2), 119–134.

Fulton, C. (2009). The pleasure principle: the power of positive affect in information seeking. In *Aslib Proceedings* (Vol. 61, pp. 245–261). Emerald Group Publishing Limited.

Ge, M., Delgado-Battenfeld, C., & Jannach, D. (2010). Beyond accuracy: evaluating recommender systems by coverage and serendipity. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 257–260). ACM.

Gibson, J. J. (1979). The theory of affordances. *The People, Place, and Space Reader*, 56–60.

Given, L. (2002). Discursive constructions in the university context: Social positioning theory and mature undergraduates' information behaviours. *New Review of Information Behaviour Research*, *3*, 127–141.

González-Ibáñez, R., & Shah, C. (2012). Investigating positive and negative affects in collaborative information seeking: A pilot study report. *Proceedings of the Association for Information Science and Technology*, *49*(1), 1–4.

González-Ibáñez, R., Shah, C., & Córdova-Rubio, N. (2011). Smile! Studying expressivity of happiness as a synergic factor in collaborative information seeking. *Proceedings of the Association for Information Science and Technology*, *48*(1), 1–10.

Guo, H (2013). A Study on Information Encountering in the Web Browser. Journal of intelligence, Vol. 32 No. 4, pp.47-50.

Hauff, C., & Houben, G.-J. (2012). Serendipitous browsing: Stumbling through wikipedia. In *Proc. 34th European Conference on IR Research (ECIR 2012)*. Citeseer.

Heinström, J. (2007). Psychological factors behind incidental information acquisition. *Library & Information Science Research*, *28*(4), 579–594.

Herlocker, J. L., Konstan, J. A., & Riedl, J. (2000). Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work* (pp. 241–250). ACM.

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, *22*(1), 5–53.

Hofstede, G. (1980). Culture and organizations. *International Studies of Management & Organization*, *10*(4), 15–41.

Hourcade, J. P. (2008). Interaction design and children. *Foundations and Trends in Human-Computer Interaction*, *1*(4), 277–392.

Iaquinta, L., Gemmis, M. de, Lops, P., Semeraro, G., Filannino, M., & Molino, P. (2008). Introducing Serendipity in a Content-Based Recommender System (pp. 168–173). IEEE. https://doi.org/10.1109/HIS.2008.25

Ji, L. J., Peng, K., & Nisbett, R. E. (2000). Culture, control, and perception of relationships in the environment. *Journal of Personality & Social Psychology*, *78*(5), 943–955.

Jiang, T., Liu, F., & Chi, Y. (2015). Online information encountering: modeling the process and influencing factors. *Journal of Documentation*, *71*(6), 1135–1157.

Kamahara, J., Asakawa, T., Shimojo, S., & Miyahara, H. (2005). A community-based recommendation system to reveal unexpected interests. In *Multimedia Modelling Conference, 2005. MMM 2005. Proceedings of the 11th International* (pp. 433–438). IEEE.

Kawamae, N. (2010). Serendipitous recommendations via innovators. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval* (pp. 218–225). ACM.

Kawamae, N., Sakano, H., & Yamada, T. (2009). Personalized recommendation based on the personal innovator degree. In *Proceedings of the third ACM conference on Recommender systems* (pp. 329–332). ACM.

Kefalidou, G., & Sharples, S. (2016). Encouraging serendipity in research: Designing technologies to support connection-making. *International Journal of Human-Computer Studies*, *89*, 1–23.

Kitayama, S., Duffy, S., Kawamura, T., & Larsen, J. T. (2003). Perceiving an object and its context in different cultures: a cultural look at new look. *Psychological Science*, *14*(3), 201–206.

Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*, *42*(5), 361.

Lamb, R., King, J. L., & Kling, R. (2003). Informational environments: Organizational contexts of online information use. *Journal of the Association for Information Science and Technology*, *54*(2), 97–114.

Lawley, J., & Tompkins, P. (2008). Maximising serendipity: the art of recognising and fostering potential. *Esitetty Tilaisuudessa The Developing Group*, *7*, 2008.

Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research methods in human-computer interaction*. Morgan Kaufmann.

Lee, K., & Lee, K. (2013). Using Experts Among Users for Novel Movie Recommendations. *Journal of Computing Science and Engineering*, *7*(1), 21–29. https://doi.org/10.5626/JCSE.2013.7.1.21

Lee, K., & Lee, K. (2015). Escaping your comfort zone: A graph-based recommender system for finding novel recommendations among relevant items. *Expert Systems with Applications*, *42*(10), 4851–4858. https://doi.org/10.1016/j.eswa.2014.07.024

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and Decision Making. *Annual Review of Psychology*, *66*(1), 799–823. https://doi.org/10.1146/annurev-psych-010213-115043

Liu, G. (2013). Amazing serendipity, SCIENCE & TECHNOLOGY INFORMATION, No. 18, pp.213

Liu, S., Cui, W., Wu, Y., & Liu, M. (2014). A survey on information visualization: recent advances and challenges. *The Visual Computer*, *30*(12), 1373–1393.

Lopatovska, I., & Arapakis, I. (2011). Theories, methods and current research on emotions in library and information science, information retrieval and human–computer interaction. *Information Processing & Management*, *47*(4), 575–592.

Makri, S., Bhuiya, J., Carthy, J., & Owusu-Bonsu, J. (2015). Observing serendipity in digital information environments. *Proceedings of the Association for Information Science and Technology*, *52*(1), 1–10.

Makri, S., & Blandford, A. (2012a). Coming across information serendipitously–part 1: a process model. *Journal of Documentation*, *68*(5), 684–705.

Makri, S., & Blandford, A. (2012b). Coming across information serendipitously–Part 2: A classification framework. *Journal of Documentation*, *68*(5), 706–724.

Makri, S., Blandford, A., Woods, M., Sharples, S., & Maxwell, D. (2014). "Making my own luck": Serendipity strategies and how to support them in digital information

environments. *Journal of the Association for Information Science and Technology*, *65*(11), 2179–2194.

Makri, S., & Warwick, C. (2010). Information for inspiration: Understanding architects' information seeking and use behaviors to inform design. *Journal of the Association for Information Science and Technology*, *61*(9), 1745–1770.

McBirnie, A. (2008). Seeking serendipity: the paradox of control. In *Aslib Proceedings* (Vol. 60, pp. 600–618). Emerald Group Publishing Limited.

McBirnie, A., & Urquhart, C. (2011). Motifs: dominant interaction patterns in event structures of serendipity. *Information Research*, *16*(3).

McCay-Peet, L., & Toms, E. (2011). Measuring the dimensions of serendipity in digital environments. *Information Research: An International Electronic Journal*, *16*(3), n3.

McCay-Peet, L., & Toms, E. G. (2010). The process of serendipity in knowledge work. In *Proceedings of the third symposium on Information interaction in context* (pp. 377–382). ACM.

McCay-Peet, L., & Toms, E. G. (2015). Investigating serendipity: How it unfolds and what may influence it. *Journal of the Association for Information Science and Technology*, *66*(7), 1463–1476.

McCay-Peet, L., Toms, E. G., & Kelloway, E. K. (2014). Development and assessment of the content validity of a scale to measure how well a digital environment facilitates serendipity. *Information Research*, *19*(3), 19–3.

McCay-Peet, L., Toms, E. G., & Kelloway, E. K. (2015). Examination of relationships among serendipity, the environment, and individual differences. *Information Processing & Management*, *51*(4), 391–412.

McNee, S. M., Riedl, J., & Konstan, J. A. (2006). Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *CHI'06 extended abstracts on Human factors in computing systems* (pp. 1097–1101). ACM.

Merton, R. (2004). *EG, The Travels and Adventures of Serendipity. A Study in Historical Semantics and the Sociology of Science*. Priceton University Press.

Mintzberg (1967). Crafting strategy, Harvard Business Review Reprints, July-August, p.66-75. Harvard Business School, Publishing Division, Boston, MA 02163.

Miwa, M., Egusa, Y., Saito, H., Takaku, M., Terai, H., & Kando, N. (2011). A method to capture information encountering embedded in exploratory Web searches. *Information Research*, *16*(3).

Mladenic, D. (1999). Text-learning and related intelligent agents: a survey. *IEEE Intelligent Systems and Their Applications*, *14*(4), 44–54.

Morris, M. W., & Peng, K. (1994). Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology*, *67*(6), 949.

Murakami, T., Mori, K., & Orihara, R. (2007). Metrics for evaluating the serendipity of recommendation lists. In *Annual Conference of the Japanese Society for Artificial Intelligence* (pp. 40–46). Springer.

Natale, V., Alzani, A., & Cicogna, P. (2003). Cognitive efficiency and circadian typologies: a diurnal study. *Personality and Individual Differences*, *35*(5), 1089–1105.

Nazemi, K., Burkhardt, D., Hoppe, D., Nazemi, M., & Kohlhammer, J. (2015). Web-based evaluation of information visualization. *Procedia Manufacturing*, *3*, 5527–5534.

Nutefall, J. E., & Ryder, P. M. (2010). The serendipitous research process. *The Journal of Academic Librarianship*, *36*(3), 228–234.

O'haire, M. E., McKenzie, S. J., Beck, A. M., & Slaughter, V. (2015). Animals may act as social buffers: Skin conductance arousal in children with autism spectrum disorder in a social context. *Developmental psychobiology*, *57*(5), 584-595.

Oku, K., & Hattori, F. (2011). Fusion-based Recommender System for Improving Serendipity. *DiveRS@ RecSys*, *816*, 19–26.

Oku, K., & Hattori, F. (2012). User evaluation of fusion-based approach for serendipity-oriented recommender system. In *Proceedings of the Workshop on Recommendation Utility Evaluation: Beyond RMSE (RUE 2012)* (pp. 39–44).

Onuma, K., Tong, H., & Faloutsos, C. (2009). TANGENT: a novel,'Surprise me', recommendation algorithm. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 657–666). ACM.

Pálsdóttir, Á. (2011). Opportunistic discovery of information by elderly Icelanders and their relatives. *Information Research*, *16*(3).

Parrott, W. G. (Ed.). (2014). *The positive side of negative emotions*. Guilford Publications.

Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In *The adaptive web* (pp. 325–341). Springer.

Pontis, S., Kefalidou, G., Blandford, A., Forth, J., Makri, S., Sharples, S., … Woods, M. (2016). Academics' responses to encountered information: Context matters. *Journal of the Association for Information Science and Technology*, *67*(8), 1883–1903.

Porat, T., & Tractinsky, N. (2012). It's a pleasure buying here: The effects of web-store design on consumers' emotions and attitudes. *Human–Computer Interaction*, *27*(3), 235–276.

Race, T. M., & Makri, S. (2016). *Accidental Information Discovery: Cultivating Serendipity in the Digital Age*. Elsevier.

Ravaja, N., Turpeinen, M., Saari, T., Puttonen, S., & Keltikangas-Järvinen, L. (2008). The psychophysiology of James Bond: Phasic emotional responses to violent video game events. *Emotion*, *8*(1), 114.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: an open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work* (pp. 175–186). ACM.

Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, *40*(3), 56–58.

Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to recommender systems handbook. In *Recommender systems handbook* (pp. 1–35). Springer.

Richardson, R. M., & Smith, S. W. (2007). The influence of high/low-context culture and power distance on choice of communication media: Students' media choice to communicate with professors in Japan and America. *International Journal of Intercultural Relations*, *31*(4), 479–501.

Robson, A., & Robinson, L. (2013). Building on models of information behaviour: linking information seeking and communication. *Journal of Documentation*, *69*(2), 169–193.

Rohfl, F. (1992). NTSYS-pc Numerical taxonomy and multivariate analysis systems. Versión 1.70. *Exeter Software, Setauket, NY*.

Roth, W. T., Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*(8), 1017–1034.

Rubin, V. L., Burkell, J., & Quan-Haase, A. (2011). Facets of serendipity in everyday chance encounters: a grounded theory approach to blog analysis. *Information Research*, *16*(3).

Saadatmand, M., & Kumpulainen, K. (2013). Content aggregation and knowledge sharing in a personal learning environment: Serendipity in open online networks. *International Journal of Emerging Technologies in Learning (IJET)*, *8*(2013), 70–77.

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web* (pp. 285–295). ACM.

Savolainen, R. (2014). Emotions as motivators for information seeking: A conceptual analysis. *Library & Information Science Research*, *36*(1), 59–65.

Scheirer, J., Fernandez, R., Klein, J., & Picard, R. W. (2002). Frustrating the user on purpose: a step toward building an affective computer. *Interacting with computers*, *14*(2), 93-118.

Schmidt, A. (2000). Implicit human computer interaction through context. *Personal Technologies*, *4*(2–3), 191–199.

Schupp, H. T., Flaisch, T., Stockburger, J., & Junghöfer, M. (2006). Emotion and attention: event-related brain potential studies. In S. Anders, G. Ende, M. Junghofer, J. Kissler, & D. Wildgruber (Eds.), *Understanding Emotions* (Vol. 156, pp. 31–51). Elsevier. https://doi.org/10.1016/S0079-6123(06)56002-9

Senoff, C. V. (1990). The discovery of [Ru (NH3) 5N2] 2+: A case of serendipity and the scientific method. *Journal of Chemical Education*, *67*(5), 368.

Shani, G., & Gunawardana, A. (2011). Evaluating recommendation systems. In *Recommender systems handbook* (pp. 257–297). Springer.

Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, *5*(1), 3–55.

Stewart, K. N., & Basic, J. (2014). Information encountering and management in information literacy instruction of undergraduate, students. *International Journal of Information Management*, *34*(2), 74–79.

Sun, T., Zhang, M., & Mei, Q. (2013). Unexpected Relevance: An Empirical Study of Serendipity in Retweets. In *ICWSM*.

Sun, X., Sharples, S., & Makri, S. (2011). A user-centred mobile diary study approach to understanding serendipity in information research. *Information Research*, *16*(3).

Sundin, O. (2002). Nurses' information seeking and use as participation in occupational communities. *New Review of Information Behaviour Research*, *3*, 187–202.

Swearingen, K., & Sinha, R. (2001). Beyond algorithms: An HCI perspective on recommender systems. In *ACM SIGIR 2001 Workshop on Recommender Systems* (Vol. 13, pp. 1–11). Citeseer.

Talja, S. (1997). Constituting "information" and "user" as research objects: a theory of knowledge formations as an alternative to the information man-theory. *Information Seeking in Context*, 67–80.

Talja, S., Keso, H., & Pietiläinen, T. (1999). The production of 'context'in information seeking research: a metatheoretical view. *Information Processing & Management*, *35*(6), 751–763.

Talja, S., Tuominen, K., & Savolainen, R. (2005). "Isms" in information science: constructivism, collectivism and constructionism. *Journal of Documentation*, *61*(1), 79–101.

Taramigkou, M., Bothos, E., Christidis, K., Apostolou, D., & Mentzas, G. (2013). Escape the bubble: guided exploration of music preferences for serendipity and novelty (pp. 335–338). ACM Press. https://doi.org/10.1145/2507157.2507223

Taylor, R. S. (2015). Question-negotiation and information seeking in libraries. *College & Research Libraries*, *76*(3), 251–267.

Tian, L.and Yu, B. (2013). Influencing Factors of Serendipity for Scholars. Information Science. Vol. 31 No. 4, pp. 69-75.

Tkalčič, M. De Carolis, B. de Gemmis, M. Odić, A. and Košir, Eds.A, Emotions and personality in personalized services. Cham: Springer International Publishing, 2016.

Toms, E. G. (2000). Understanding and facilitating the browsing of electronic text. *International Journal of Human-Computer Studies*, *52*(3), 423–452.

van Reekum, C., Johnstone, T., Banse, R., Etter, A., Wehrle, T., & Scherer, K. (2004). Psychophysiological responses to appraisal dimensions in a computer game. *Cognition and Emotion*, *18*(5), 663–688.

Wang, C., Zhou, Z., Jin, X.-L., Fang, Y., & Lee, M. K. (2017). The influence of affective cues on positive emotion in predicting instant information sharing on microblogs: Gender as a moderator. *Information Processing & Management*, *53*(3), 721–734.

Wang, M., Kawamura, T., Sei, Y., Nakagawa, H., Tahara, Y., & Ohsuga, A. (2013). Context-aware music recommendation with serendipity using semantic relations. In *Joint International Semantic Technology Conference* (pp. 17–32). Springer.

Wang, Q., & Conway, M. A. (2004). The stories we keep: autobiographical memory in American and Chinese middle-aged adults. *Journal of Personality*, *72*(5), 911–938.

Williamson, K. (1998). Discovered by chance: The role of incidental information acquisition in an ecological model of information use. *Library & Information Science Research*, *20*(1), 23–40.

Wilson, T. D. (1981). On user studies and information needs. *Journal of Documentation*, *37*(1), 3–15.

Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation*, *55*(3), 249–270.

Wu, K.-C. (2015). Affective surfing in the visualized interface of a digital library for children. *Information Processing & Management*, *51*(4), 373–390.

Yadamsuren, B., & Heinström, J. (2011). Emotional reactions to incidental exposure to online news. *Information Research*, *16*(3), 16–3.

Yamaba, H., Tanoue, M., Takatsuka, K., Okazaki, N., & Tomita, S. (2013). On a serendipity-oriented recommender system based on folksonomy. *Artificial Life and Robotics*, *18*(1–2), 89–94. https://doi.org/10.1007/s10015-013-0103-3

Yeh, N.-C. (2006). A framework for understanding culture and its relationship to information behaviour: Taiwanese aborigines' information behaviour. *Information Research*, *12*(2), 11.

Zaslow, J. (2002). If TiVo thinks you are gay, here's how to set it straight. *Wall Street Journal*, *26*.

Zhang, Y., Kang, J., & Jin, H. (2015). An experimental study on the restorative experience of typical urban soundscape based on EDA. *Italy: Iiav. Org*.

Zhou, P. and Huang, Ch (2014). Research on Information Encountering Based on Meta-ethnography, Library and Information Service, Vol. 58, No.14, pp 115-120.

Zhou, X., Sun, X., Wang, Q., & Sharples, S. (2018). A context-based study of serendipity in information research among Chinese scholars. *Journal of Documentation*.

Appendix 1 User Study 1 —— Interview Protocol

Q1: How long have you been in the research of serendipity study?

Q2: What are the challenges or problems of applying conventional research methods (e.g. interview, survey, focus group) in studying serendipity?

Q3: Imagine this scenario:

Ann is concerning about an important literature of her research; however, she couldn't get access with the paper. One day, when Ann is searching online in her spare time, a website link jumped into her eyes which said it can freely download academic papers. Curiously, Ann opened the link and tried to search the paper in the website. Although she failed to download the paper immediately, she found the author's information in this website. She then tried to send an email to the author inside the website and asked if the author could send her the paper. To her surprise, she received the author's reply inside the website with the paper as the attachment the next day.

In this scenario, what are the interesting aspects you may look into and how will you collect data relating to these aspects?

Q4: What research method have you used before to conduct serendipitous research?

Q5: What is your user population in your previous study?

Q6: What is your research context in your previous study?

Q7: What are the advantages of applying your research methods in studying serendipity?

Q8: What are the disadvantages of applying your research methods in studying serendipity?

Q9: Having discussed the pros and cons of these research methods, how do you suggest balancing the methods to conduct a good serendipitous research?

Appendix 2 User Study 2 —— Interview Protocol

Part 1: Understanding of Serendipity

(1) What is your understanding of serendipity now that the study has concluded?
(2) When comparing your current understanding of serendipity to your understanding before the study, do you think are there any differences? If yes, why?

Part 2: For a Detailed Serendipity Case

(1) When did the case happen?
(2) Where did the serendipity happen?
(3) Could you describe the case in more detail?
(4) Why would you consider it as serendipity?
(5) What was your socialisation context in this case?
(6) How did you deal with the serendipitous information? What did you do when you encountered the serendipity?
(7) What was your emotion after you encountered the serendipity?

Part 3: Questions about the Research Method Employed

(1) What do you think about the reminder information that I sent to you every day?
(2) Do you have any suggestions or opinions on such reminders?
(3) Why did you send your case in text (or picture)?
(4) What do you think about the designed interface of the application?
(5) What was your experience of the study? Do you have any suggestions or opinions about the study or the research method?

*Note: The language used across the study was Chinese, including all the contents presented in the appendices. We have translated everything into English for this paper.

Appendix 3 User Study 2 ——A Description for the Content of Figure 5-2

Figure 5-2-a: a description of the five notes on the left hand side of the picture:

a) *Time*: represents the time when the participant comes across serendipity;

b) *Location:* represents the location where the participant comes across serendipity (e.g. in the classroom, in the dormitory, on the street, in the library, etc.);

c) *Activity:* represents the behaviour when serendipity happens (e.g. searching the Internet, chatting with others, reading a book/literature, listening to music, watching an educational TV programme, etc.);

d) *Emotion:* represents the emotion experienced when serendipity happens (e.g. happiness, surprise, interest, sadness, stress, etc.);

e) *Impact:* represents the influence and follow-up behaviour when serendipity happens (e.g. store the information, use it immediately, ignore it and do nothing, etc.)

Figure 5-2-b: illustrations of the user input sections:

As can be seen from the image, there are different input sections for the participant to record and send the encountered serendipity throughout the study, including voice, text, picture and video.

Figure 5-2-c: The meaning of the sent messages (daily reminder), taking the first message as an example:

"I have received eleven messages today, ten days left for the study" indicates that the researchers have successfully received eleven serendipity messages from all the participants on that day, and there are ten days left before the study finishes.

Appendix 4 User Study 3 —— Interview Protocol

Q1: How do you think about the game? Is it fun?

Q2: What do you think about the study? Is it acceptable to you?

Q3: Do you think the study have impact on your emotion?

Q4: What do you think about the provided information? Why?

Q5: (With the filled questionnaire) Why you give this score on this dimension?

Q6: What do you think about the way to get access of the information in such a “game” context? Why?

Q7: Do you have any suggestions on the study?

Appendix 5 List of Publications

(Chinese) Patent for Invention

[1] Name: A new recommendation method.

Inventor: Xu Sun, Xiaosong Zhou

Patent number: CN106874463A

[2] Name: An Android-based patent image recognition method

Inventor: Xu Sun, Xiaosong Zhou, Fangyan Tang, Qingfeng Wang, Ye Zeng, Die Hu

Patent number: 107622278A

Publications

[1] Xiaosong Zhou, Zhan Xu, Xu Sun & Qingfeng Wang. (2017, July). A New Information Theory-based Serendipitous Algorithm Design. In International Conference on Human Interface and the Management of Information (pp. 314-327). Springer, Cham.

[2] Xiaosong Zhou, Xu Sun, Zhan Xu. (2016). A Methodology Based Empirical Study of Serendipity in Digital System Design. SEADE workshop Paper presented in ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR). Chapel Hill, North Carolina, USA.

[3] Xiaosong Zhou, Xu Sun, & Sharples Sarah. (2014). Methods to Study Serendipity. User friendly 2014. Wuxi, China.

[4] Xiaosong Zhou, Xu Sun, Qingfeng Wang & Sharples Sarah. A Context-based Study on Serendipity in Information Research among Chinese Scholars. *Journal of Documentation*.

[5] Xiaosong Zhou, Xu Sun, Zhan Xu & Qingfeng Wang. The Impact of Emotions on Experiencing Serendipity in a Controlled Laboratory. *Information Processing and Management*. (Under Review)

[6] Xiaosong Zhou, Xu Sun & Sharples Sarah. A Review of Methods on Serendipity Studies in the Context of Information Research. *Information Research*. (Ready to submit)

[7] Xiaosong Zhou, Xu Sun, Xiangjun Peng, Hongzhi Zhang & Qingfeng Wang. A New Information Theory-based Serendipitous Algorithm in Recommendation Systems. *International Journal of Human-Computer Studies*. (Ready to submit)