國立臺灣大學管理學院資訊管理學研究所

碩士論文

Graduate Institute of Information Management
College of Management
National Taiwan University

Master Thesis

推薦系統中意外發現之觸發及其對使用者滿意度影響

On the Approaches to Triggering Serendipity in Recommender Systems and Their Impacts to User Satisfaction

林佑宣
Yu-Hsuan Lin

指導教授：吳玲玲 博士
　　　　　莊裕澤 博士

Advisor: Ling-Ling Wu, Ph.D.
　　　　Yuh-Jzer Joung, Ph.D.

中華民國 101 年 8 月

August, 2012

謝辭

論文的完成，象徵研究所求學階段的結束。在工作將近一年後，對於這期間所發生的種種我都充滿了感激的心情。首先要感謝我的兩位指導老師，吳玲玲老師以及莊裕澤老師，無論對於研究亦或是處世態度，都在在給予學生許多寶貴的建議。藉由兩位老師在研究方面的細心指導、深入問題核心的評論，指引學生正確的方向，才讓我的研究能夠順利完成。三年的研究所生涯，讓我理解到研究的本質並非只是為了完成最後的論文，更重要的是在整個研究過程中學習做研究的態度與方法，更甚是人與人之間相處及互動的關係。

對於在百忙之中撥冗，並接受時間安排唐突的口試的系上的兩位教授：許瑋元老師及盧信銘老師，也非常感謝獲得老師們的首肯，願意擔任學生的口試委員。兩位老師在口試時，對於論文提出諸多鞭辟入裡的建議，並點出許多與研究或論文撰寫相關的問題，令我受用無窮。

研究生活的最後一年，對我也是人生中重要的一年。我開始了第一份工作，並在工作的過程中磨練自我的能力。感謝公司主管信銘經理及宜霈襄理對我的支持，讓我得以在工作外兼顧研究論文。此外，對於曾在研究上給予許多經驗談的宗霖學長及小金、在研究所一起共患難的書紳、阿任，同實驗室一起奮鬥兩年的美鬧、小殼、惠菁和子龍，我都感謝一路上他們的陪伴。最後，也謝謝小旭平日的開導，讓我在情緒低落時給予我正面的關懷。

最後，對於在研究所三年期間，願意一直在背後支持與鼓勵我的家人，我也充滿感謝之情。雖然就學期間我遇到了許多挫折，但回家後母親溫暖的包容都能夠讓我忘卻煩惱。除此之外，對於遠在美國求學的妹妹，以及在大陸工作的父親所給予我的支持，也是幫助我完成學業的推手。因此，最後我要再次對我的家人說聲感謝！

林佑宣

民國 101 年十一月

中文摘要

本研究主要針對兩種主要的推薦系統策略：協同過濾及內容導向，並在推薦過程中導入隨機性與降低準確性的方法，藉以觀察隨機性或準確性的降低對於刺激推薦商品中意外驚喜之發生，及對傳統用於評估推薦結果品質的各項指標如滿意度、購買意願等之影響。本研究採取實驗法以驗證假設，受測者隨機分配各特定專為實驗設計之推薦系統後，於一個虛擬電影租賃網站進行購買決策行為。待實驗結束，受測者以填寫問卷的方式回報其感興趣程度、滿意度與購買意願等指標。實驗結果證實意外驚喜的提升與其他各項指標間存在互換關係。除此之外，協同過濾型的推薦系統配合降低準確性的作法，是最適合刺激意外驚喜發生的推薦系統策略；這樣的組合能夠在不犧牲現有推薦品質的情況下提高意外驚喜出現的比例。最後，針對推薦的候選商品加上特定過濾條件如較高商品評價之門檻，將有助於減緩上述意外驚喜與其他衡量指標間之互換關係。本研究的結果對於推薦系統中意外驚喜的相關研究有重要意涵。

關鍵字：推薦系統、意外驚喜、滿意度、購買意願、隨機性、降低準確性

ABSTRACT

This study focuses on two main recommender paradigms: collaborative-filtering and content-based, and introduces the "Role of chance" approach and the "Anomalies and exceptions" approach. The above two approaches are integrated in this study to form a theoretical model that examines their effects on triggering serendipity and the subsequent effects on several metrics such as user satisfaction and willingness to pay. An experiment was conducted to test the model. Participants were grouped by each recommender conditions and were asked to make a purchase at a simulated online retailer. After the experiment, participants were asked to complete a survey to report their interest, satisfactory and willingness to pay levels. Results indicate that there might be a trade-off relationship between serendipity and other metrics. In addition, collaborative-filtering recommenders which adopted the "Anomalies and exceptions" approach seem to be the most suitable combination to introduce serendipity. Finally, setting a threshold to filter products among recommendation candidates such as high rating would ease the trade-off. Our findings have major implications for the ongoing research on serendipity of recommendations.

Keywords: Recommendation systems; Serendipity; User satisfaction; Willingness to Pay; Role of Chance; Anomalies and Exceptions

Table of Contents

Table Index

1. Introduction

Over the last decade, followed the Internet growth, the number of products available have seen an extraordinary increase in the online market (Brynjolfsson et al. 2006; Clemons et al. 2006), and consumers have taken to these expanded offering. According to Resnick and Varian (1997), although the Internet has allowed people to publish information easily, people are faced with a problem called “information overload” subsequently. Such information availability and depth may attributed to customers engaging in more online consumptions (Van den Poel & Leunis, 1999; Zellweger, 1997), but too much information have made it too difficult for customers to find desirable products from the endless of alternatives available (Fleder & Hosanager, 2009). Recommender systems are considered one solution to help customers to solve this problem.

Recommender systems have shown a great ability to help users find interesting and relevant products from large catalogues of items, in addition to indicating those that should be filtered out. According to Resnick and Varian (1997), the main objective of recommender systems is to help users find preferable products among a potential large product pool. Customized recommendations to a particular user could be provided by tracking user data such as purchase history, product ratings or user profiles (Burke, 2002). In other words, recommender systems made an effort to analyze users' characteristics and past behavior to predict the products that best suit users' preference. Nowadays, recommender systems are changing from novelties used by a few online retailers, to serious applications widely adopted by E-Commerce websites.

Generally, recommender systems can be classified into two main paradigms, base on how recommendations are made: content-based systems and collaborative filtering systems (Balabanovic & Shoham, 1997). The content-based approach is based on user's preference and the collaborative filtering approach is based on k-nearest neighbors. Most of the previous researches of recommender systems have focused on improving the accuracy or efficiency of recommender systems, to precisely predict the purchase probability of users and recommend products which met user's need. That is to say, traditionally the success of recommender systems is evaluated by predicting accuracy of recommendations. Dozens of quantitative metrics such as Mean Absolute Error (Herlocker, Konstan, Borchers, & Riedl, 1999) and Precision/Recall (Sarwar, Karypis, Konstan, & Riedl, 2000) are developed to archive this goal. This viewpoint may misguide, since such techniques present the problem that they recommend many items that the user already known. According to McNee (2006), the most accurate recommendations are sometimes not the most useful recommendations to users. Previous researches have also shown that user satisfaction does not always correlate with high recommender accuracy (McNee, 2002; Ziegler, 2005).

A “correct” recommendation isn’t necessarily be good, only if we consider accuracy alone. On the other hand, when we consider users’ satisfaction, the lack of discovery or diversity may lead to negative effects (Ziegler, 2005). For example, an accuracy movie recommender suggests recently popular movies to customers on online movie rental website. Though being highly accurate, note that almost everyone likes hit movies. Hence, the recommendations appear far too obvious and of little help to the user. Due to the above reasons, an accurate or familiar recommendation but the user has already known is not really helpful, caused by such recommendations do not give users any new information. Previous Studies suggest that beyond accuracy there is other metrics to evaluate recommenders, such as coverage, serendipity, and transparency (Herlocker, et al., 2004; Tintarev and Masthoff, 2007). In this study, we will focus on evaluating two common approaches of recommenders: content-based and collaborative filtering.

Previous researches suggest that collaborative filtering recommender has seen a great potential to provide more unexpected or different recommendations because of the content-based approach suffers from over-specialization (Balabanovic & Shoham, 1997). When the system can only recommend items that score highly against a user’s profile, the user is limited to being recommended items similar to those already rated. This shortcoming is called serendipity problem. Over-specialized systems are hard to make serendipitous discoveries happen, according to Gup's theory (Gup, 1997). However, collaborative filtering has also been shown to over-specialize in some cases (Ziegler et al., 2005). The problem is that collaborative filtering systems tend to focus on what is commonly known and popular - which most of the items in the generated recommendation are items that the user has heard about or items that the user would have experienced eventually because of the “blockbuster culture” (Fleder & Hosanager, 2009). Therefore, users easily get tired of the recommendations if there is nothing new or something serendipitous.

Whereas the inherent over-specialize problems of content-based and collaborative filtering recommender systems, we conduct a common way to trigger serendipities happen both on content-based and collaborative filtering recommender systems in this study. According to Toms (2000), we introduce the “Role of chance” approach (implemented via a random information node generator) and the “Anomalies and exceptions” approach (partially implemented via poor similarity measures) to our experiment trying to provide the user with new entry points to the items in the recommender systems. The basic assumption is that one recommendation cannot be serendipity if the user already knows what is recommended to her, because a serendipitous happening is by definition something new. Thus, the first purpose of this study is to adopt the above two approaches and examine their effects on serendipity generated by content-based and collaborative filtering recommenders.

A serendipitous recommendation is an unexpected and desirable recommendation, helps the user find an interesting item he might not have otherwise discovered (Herlocker, et al., 2004, Schafer, et al., 2007). Stated informally, serendipity means "tell me something I don't already know." It is not always the case that the users are interested in what is similar to what they has already known. Swearingen and Sinha (2001) also indicated that from a user's perspective, an effective recommender system should point users towards new, not-yet-experienced items. As the above mentioned, Serendipity of recommendations may affect the user satisfaction toward recommender systems, and affect the acceptance of the user toward recommendations. According to Delon and Mclean (1992, 2003), user satisfaction is a key factor of measuring Information System success, and is one of the most important consumer reactions in B2C online market (Cheung & Lee, 2005). Moreover, willingness to pay (Homburg, Koschate and Hoyer, 2005), a particular important behavioral outcome has been linked to user satisfaction. Whereas satisfaction is considered a cognitive and affective outcome (Oliver, 1997), willingness to pay is a behavioral consequence that directly related to firm profitability. Owing to the above reasons, the second purpose of this study is to directly examine the impact of several recommender systems (adopted Role of chance/Anomalies and exceptions approaches or not) on user satisfaction, willingness to pay and serendipity, and compare their differences.

The remaining study is structured as followed: Chapter 2 gives a thorough review of the literature on topics regarding serendipity, recommendation strategies, user satisfaction, as well as willingness to pay. In addition, according to the review, the hypotheses of this study are stated. In Chapter 3, the research methodology of this study is specified, including experimental design, experimental procedure, participants of the experiment and measurements of the independent and dependent variables. Empirical results are presented in Chapter 4, with discussions, managerial implications, limitations as well as directions for future researches presented in Chapter 5.

2. Literature View

2.1. Serendipity

Traditionally, the success of recommender systems is evaluated by predicting accuracy of recommendations, such as the Mean Absolute Error (Herlock et al., 1999) and Precision/Recall (Sarwar et al., 2000). For instance, Netflix provide one million dollars award for who substantially improve the accuracy of their collaborative filtering algorithm. Recommendations based upon this traditional accuracy metric are not the most useful to users (McNee, 2006). In other words, an accurate but known recommendation is an item that the user might already aware of. When analyzing recommender systems, we need to consider the "non-obviousness" of the recommendation.

Recently, several researchers began to investigate the aspect of serendipity in the context of recommender systems. Iaquinta et al. (2008) found that serendipity is mostly related to the quality of recommendations and largely depends on subjective characteristics. Various definitions are proposed for this concept in the recommender system domain. For example, according to Herlocker et al. (2004) and Schafer et al. (2007), serendipity is defined as an unexpected and desirable recommendation which helps the user find an interesting item he might not have otherwise discovered. Based on this definition, Ge, Delgado-Battenfeld and Jannach (2010) have presented two important aspects related to serendipity. First, a serendipitous product should be undiscovered and unexpected by the user; secondly, the product should also be interesting, relevant and useful to the user. Similar definitions are also found in other researches. For instance, Shani and Gunawardana (2009) considered serendipity as a measure of how surprising and successful the recommendations are. On the other hand, McNee et al. (2006) defined serendipity as the experience of the user who received an unexpected and fortuitous recommendation.

In order to facilitate our experiment, we adopt the two aspects of serendipity by Ge et al. (2010). We define serendipitous recommendations must be something new that the user not already aware of and also be attractive to the user. One of the main purposes of this study is to examine the impact of conducting Toms' approaches of recommender systems on serendipity. According to Iaquinta et al. (2008), serendipity is largely depends on subjective characteristics. Whereas serendipity seems difficult to directly evaluate, we must rely on feedback from users. A questionnaire was adopted to explore the user perception of serendipity. For each of the recommended products and the final products that the consumer has chosen to purchase, we asked the user about these products he/she had ever known before and his/her interest levels toward these products. Products which garnered higher interest levels and were previously unknown to the user would be considered as serendipities.

2.2. Recommender System

In this study, we adopt two of the most popular recommender strategies and investigate their impacts on user perception of serendipity. The two strategies are content-based and collaborative filtering systems respectively.

According to Bilgic and Mooney (2005), Collaborative filtering (CF) systems recommend items by matching a user's tastes to those of other users of the system. CF systems make automatic predictions (filtering) about the interests of a user by collecting taste information from other users that are highly correlated (collaborating) (Good, Schafer et al. 1999). It then recommends items by predicting the purchase probability for the items that a user has not purchased but the neighbors has. For example, if Tom and John have watched a lot of identical movies, and Tom watched a new movie while John hasn't, the system will most likely recommend this new movie to John. To sum up, CF systems will recommend the most popular products among groups of similar users.

On the other hand, Content-based (CB) systems recommend items based on items' own attributes rather than other users' ratings (Bilgic & Mooney, 2005). CB systems analyze item descriptions (content) to identify items that are of particular interest to the user (personal preference) and return highly correlated results (Pazzani & Billsus, 2007). For CB systems, in order to accurately predict, user profile must first be acquired, either through explicit methods (i.e. a user preference survey) or implicit methods (i.e. analyzing user purchase history). CB systems are then compare user's preference in the user profile with product's attributes and recommend products which most fit the user. For example, if Tom likes watching action movies and hates horror movies, CB systems will most likely recommend action movies rather than horror movies to him. To sum up, CB systems will recommend the most similar products according to user's preference.

Previous studies have pointed out that both content-based and collaborative filtering recommender systems suffer from over-specialization problems (Balabanovic & Shoham, 1997; Ziegler et al., 2005). Further, Gup (1997) has also suggested that over-specialized systems are hard to make serendipitous discoveries happen. In order to introduce serendipity in the recommendation process in an operational way, Toms (2000) suggested four strategies among different approaches which have been proposed (Bawden, 1986), from simplistic to more complex ones: (1) Role of chance or "blind luck", implemented via a random information node generator. (2) Pasteur principle ("chance favors the prepared mind"), implemented via a user profile. (3) Anomalies and exceptions, partially implemented via poor similarity measures. (4) Reasoning by analogy, implemented via similar patterns or hidden relationships. For example, Hijikata, Shimizu, and Nishida (2009) calculated the probability of known items using the information about known or unknowns given explicitly by user for improving serendipity. Furthermore, Kawamae, Sakano, and Yamada (2009) suggested an algorithm for recommending novel items based on the assumption that users follow the earlier adopters who have demonstrated similar preferences but purchased items earlier. We found out that previous researches mostly focus on the "Pasteur principle" approach and the "Reasoning by analogy" approach, and rely on collecting lots of user's historical data. Thus, in our study we implemented the "Role of chance" strategy and the "Anomalies and exceptions" strategy to handle the over-specialization problems which mentioned before and provide serendipitous recommendations alongside classical recommender systems.

2.3. User Satisfaction and Willingness to pay

According to Delone and Mclean (1992, 2003), user Satisfaction is a key measure of computer system success. In general, recommenders can also be considered as computer systems, since they both rely on information quality and system quality to succeed (DeLone & McLean, 2003; McKinney, Yoon and Zahedi, 2002). However, a few researchers have argued that the baseline measure of a recommender system success should be user satisfaction (Herlocker et al., 2004, Ziegler, 2005). Previous researches evaluated the impact of recommenders on user satisfaction have adopted different dimensions as measures of satisfaction. For example, Swearingern and Sinha

(2001) reported that perceived usefulness, novelty and usability are highly correlated with user satisfaction. Their results also indicate that effective recommenders inspire trust in the system and point users towards new, not-yet-experienced items.

In this study, we would like to examine the impact and difference between several recommenders (adopted Role of chance/Anomalies and exceptions approaches or not) on user satisfaction and willingness to pay in the E-Commerce context. In order to evaluate user "reaction" toward our designed recommender systems, we adopted explicit (ask) and implicit (observe) measures for user evaluation (Herlocker, et al., 2004). A basic distinction between evaluations is that explicitly measures ask users about their reactions to a system and typically employ survey and interview methods. In contrast, implicitly measures observe user behavior and usually consist of logging user behavior, then subjecting it to various sorts of analyses.

For implicit measures, we simply measured users' satisfactory levels by the average number of recommended products purchased/put into shopping cart list and the percentage of purchased products/products in the Shopping Cart which are recommended. For explicit measures, we asked users about their overall satisfaction with recommendations. We didn't adopt product satisfactory measures for each of the recommended products and the final chosen products because of the duplicate of interest levels, the similar metric used to evaluate serendipity.

In contrast to user satisfaction, willingness to pay was examined on a product level. Willingness to pay has been found to have a positive relationship with consumer satisfaction (Homburg et. al., 2005). However, satisfaction or high level of interest alone does not guarantee purchase, since a consumer might be even more satisfied with their own choices compared to that of recommended items. Due to willingness to pay also holds importance and needs to be separately examined, the willingness to pay was also measured for each of the recommended products.

2.4. Hypotheses Development

According to Ge et al. (2010), a serendipitous recommendation must be novelty and also be attractive to the user. On the other hand, serendipity is an unexpected recommendation base on user preference. Balabanovic and Shoham (1997) have suggested that content-based system suffers from over-specialization because of the system can only recommend items that score highly against a user's profile. However, Ziegler et al. (2005) have pointed out that collaborative filtering has also been shown to over-specialize. That means collaborative filtering systems tend to focus on what is commonly known and popular - which most of the recommendations are items that the user has heard about or items that the user would have experienced. Although collaborative filtering system may provide relevant and interesting recommendations to the user, it is hard to make the user be surprise. The basic concept of serendipity is that it should be novelty and interesting. Whereas over-specialized systems are hard to make serendipities happen (Gup, 1997), we implemented the “Role of chance” approach and the “Anomalies and exceptions” approach mentioned in the literature review to introduce serendipity in the recommendation process.

In the “Role of chance” approach, we implement via random recommendations, namely each item in our dataset has equal probability to be recommended. This approach triggers serendipity rely on blind luck and focuses on the unexpected aspect of serendipity. According to Bawden (1986), a fortuitous chance observation is not necessarily restricted to closely-focused ’relevant’ information; In contrast, the information may be apparently divorced from the problem at hand. One particularly relevant form of chance is the accidental finding of something of interest while looking for something entirely different. Furthermore, in the “Anomalies and exceptions” approach, we implement via less precise algorithm of recommender systems which set a similarity threshold and then recommend items beyond the threshold with equal probability. For the “Anomalies and exceptions” approach, we want to broaden coverage of recommended items and provide relevant and interesting information to the user.

The serendipity problem, which has also been studied in other domains, is often addressed by introducing some randomness (Adomavicius & Tuzhilin, 2005). In certain cases, products should not be recommended if they are too similar to which the user has already known. To avoid confusion, in this study we call recommender systems which implemented the “Role of chance” strategy as “random recommenders”; in the same way, recommender systems which implemented the “Anomalies and exceptions” strategy are called “recommenders with less precision”. In addition, classical recommenders (including content-based or collaborative filtering) are named to be “classic recommenders” in our study. According to the above discussion, our first set of hypotheses is:

H1: Consumers aided by random recommenders can perceive more serendipity of recommended products than consumers aided by classic recommenders.

H2: Consumers aided by recommenders with less precision can perceive more serendipity of recommended products than consumers aided by classic recommenders.

Further, Herlocker et al. (2004) and Schafer et al. (2007) have pointed out that serendipity means an unexpected and desirable recommendation, that is to say, user might be interested with and not have otherwise discovered. Recommenders which suffer from over- specialization lead to result sets with items that are too similar to one another, thus reducing the diversity of results and limiting user choices (Adomavicius & Tuzhilin, 2005). Such a homogeneous set of alternatives could easily turn off the user and lower his/her interest in the site over time. Thus, a serendipitous recommendation would be preferable to the user. Refer to this aspect and the above hypotheses, we also hypothesized that:

H3: Consumers aided by random recommenders will purchase more serendipitous products than consumers aided by classic recommenders.

H4: Consumers aided by recommenders with less precision will purchase more serendipitous products than consumers aided by classic recommenders.

Secondly, we consider the effect that random recommenders and recommenders with less precision have on interest, user satisfaction and willingness to pay. According to Ge et al. (2010), for items the user has already heard a lot about (think of currently top-selling, generally-liked items) even an accurate recommendation would not be too meaningful. Granovetter (1973) has suggested that the strength of weak ties enable reaching populations and audience that are not accessible via strong ties. Applying this concept to our study, accurate recommendations can be considered as strong ties and random or less accuracy recommendations can be considered as weak ties. Because of accurate recommendations might be similar to something the user has already aware of, random or less accuracy recommendations can provide an opportunity for the user to explore something he/she has not known yet.

As mentioned before, Swearingern and Sinha (2001) examined how usefulness, novelty and usability are related to user satisfaction and reported that they are significantly correlated. Previous researches have also shown that user satisfaction does not always correlate with high recommender accuracy (McNee, 2002; Ziegler, 2005). When we consider users' satisfaction, the lack of discovery or diversity may lead to negative outcomes (Ziegler, 2005).

Although both random recommenders and recommenders with less precision could broaden user's vision by providing additional information instead of something user already knows, there still existed a qualitative distinction between them. In our study, we define serendipity must be recommendations that the user not already known and also be attractive to the user. Stated concisely, a serendipitous recommendation is not only an accurate but also a novelty recommendation. The basic difference between random recommenders and recommenders with less precision is the former not consider the relevance and only rely on "blind luck". As a result, the overall recommendation quality of random recommenders would not be high. In contrast, recommenders with less precision can not only provide new information to the user but also relevant recommendations and take account of quality. Thus, consumers will be more satisfy with recommenders with less precision and tend to accept the recommendations. Hence, our second set of hypotheses is:

H5a: Consumers aided by random recommenders are less interested with recommended products than consumers aided by classic recommenders.

H5b: Consumers aided by random recommenders are less satisfied with recommended products than consumers aided by classic recommenders.

H5c: Consumers aided by random recommenders are less willing to rent with recommended products than consumers aided by classic recommenders.

H5d: Consumers aided by random recommenders will accept less recommended products than consumers aided by classic recommenders.

H6a: Consumers aided by recommenders with less precision are more interested with recommended products than consumers aided by classic recommenders.

H6b: Consumers aided by recommenders with less precision are more satisfied with recommended products than consumers aided by classic recommenders.

H6c: Consumers aided by recommenders with less precision are more willing to *rent with recommended products than consumers aided by classic recommenders.*

H6d: Consumers aided by recommenders with less precision will accept more recommended products than consumers aided by classic recommenders.

In addition, we thought that there would be a similarity pattern with final chosen products which reflect on the consumption behavior according to the above discussion. Therefore, our third set of hypotheses is:

H7a: Consumers aided by random recommenders are less interested with final chosen products than consumers aided by classic recommenders.

H7b: Final chosen products of consumers aided by random recommenders will contain less recommended products than consumers aided by classic recommenders.

H8a: Consumers aided by recommenders with less precision are more interested with final chosen products than consumers aided by classic recommenders.

H8b: Final chosen products of consumers aided by recommenders with less precision will contain more recommended products than consumers aided by classic recommenders.

3. Research Methodology

3.1. Experiment Design

In this study, the only one independent variable was recommender strategy (browsing, random, content-based, collaborative filtering, content-based with less precision, collaborative filtering with less precision, content-based with high rating less precision, collaborative filtering with high rating less precision), which a different group of subjects is used for each level of the variable. Therefore, we adopted an eight between-subjects factorial design for our experiment. The groups are shown in below:

Recommender System			
NO	CB	CBL	CBHL
RAN	CF	CFL	CFHL

Table 3-1: Experiment Groups

In order to illustrate the independent variable clearly, we classified the different recommender strategies into 3 categories based on their characteristic: (1) Classic recommender: collaborative filtering (CF) and content-based (CB). (2) Recommender with less precision: content-based with less precision (CBL), collaborative filtering with less precision (CFL), content-based with high rating less precision (CBHL) and collaborative filtering with high rating less precision (CFHL). (3) Random recommender (RAN): used to test the "Role of chance" approach and set as a baseline to ensure all the other recommenders were actually effective. Among these recommender strategies, "browsing" (NO) which meant no recommender was considered to be a control group for evaluating serendipity and interest level and not related to our hypotheses.

3.1.1 Choice of Product

The main objective of this study is to examine the impact different approaches of recommenders have on serendipity, user satisfaction and willingness to pay in a B2C online retail environment. To verify our hypotheses, we select DVD movies as our choice of product and online DVD rental as our main experiment context. There are numerous reasons behind such a choice. First, the DVD movie catalog is diverse enough to meet the needs of various different kinds of consumers. Secondly, deep and detailed movie data is readily available online, allowing our recommenders to have enough data to trigger serendipitous encounters. Finally, the online DVD rental service is still in its infant stages, eliminating any possible biases that our participants may exhibit while doing the experiment.

3.1.2 Data Translation

In this study, we used the movie dataset collected by Wu, Joung and Lee (2011) for recommendation calculation and movie catalog for our simulated online retailer. These dataset is based on the Netflix Prize dataset which includes 100 million ratings (rating from 1 to 5) from roughly 480 thousand users on 17,770 movies (released between 1927 and 2005). Next, in order to facilitate recommendation calculation for both collaborative filtering recommenders and content-based recommenders, the users' preference vectors and movies' attributes vectors with matching dimensions need to be elicited from the data. As the Netflix Prize dataset lacks any attributes for the movies other than release year, Wu, Joung and Lee obtained another set of movie data from IMDB and matched the two datasets to obtain sufficient movie attributes. The IMDB dataset contains roughly 400 thousand movie titles along with their release years and respective genres. There were a total of 23 genres (Action, Adventure, Animation, Comedy, Crime, Documentary, Drama, Fantasy, Family, Film-Noir, Horror, Musical, Mystery, Romance, Sci-Fi, Short, Thriller, War, Western, History, Sport, Biography, Music), and each movie can belong to more than one genre.

After matching the movie titles, Wu, Joung and Lee modeled the attributes of each movie by the genres it belongs. Let x.Attr denote the attribute vector of movie x, and x.Attr[i], a Boolean value, representing if movie x belongs to genre i. With this information available, along with the 100 million ratings, the preference vectors of all 480 thousand Netflix users were also obtained. Let u.Pref denote the preference vector of user u, and u.Pref[i] representing user u's preference level towards genre i. Suppose user u has rated $x_1,\ldots,x_n$ movies, with $r_1,\ldots,r_n$ being the ratings, then u.Pref[i] is calculated with the following equation:

$$u.\mathrm{Pref}[i] = \frac{\sum_{j=1}^{\mathrm{n}} x_j.\mathrm{Attr}[i] * r_i}{n}$$

For example, according to Table 3-2, the user has the following preference vector of each genre: u.pref = [8/3, 1, 4/3, 3].

	Action	**Crime**	**Drama**	**History**	**Rating**
Movie 1	1	0	0	1	5
Movie 2	0	0	1	1	4
Movie 3	1	1	0	0	3

Table 3-2: Users' Preference Vector

Further, in consideration of resembling to an online DVD retailer, additional information such as movie posters and movie descriptions are essential. To meet the above requirement, Wu, Joung and Lee manually retrieved movie information from AtMovies, a commercial movie website based in Taiwan that contains a catalog of roughly 18,000 movies. They retrieved the Mandarin title, director, actor list, movie length, movie poster and movie description information of all 18,000 movies from the website, and matched the movies with the Netflix+IMDB hybrid dataset.

A major drop-off in the number of correctly matched movies between the datasets happened due to major differences between them. First, the AtMovies dataset consisted of movies from both the Western and Eastern hemispheres, whereas the Netflix+IMDB hybrid dataset consisted of mostly Western movies. Second, a portion of movies in the AtMovies dataset only contained the English title and the release year of the movie and was excluded. Finally, the Netflix+IMDB dataset contained multiple versions of the same movie for some titles which only counted as one movie. Due to the above reasons, Wu, Joung and Lee obtained a union of 4,021 movies after the matching, which used to form the basis of the experiment in this study. Although the movie catalog have been shrunk from 17,770 to 4,021, the dataset still contained more than 70 million ratings by roughly 380 thousand users, indicating that the 4,021 movies were mostly relevant movies that have been extensively rated by Netflix users.

3.1.3 Recommendation Strategy

Based on the dataset mentioned in the previous section, a movie x's attribute could be modeled by the genres it belongs and then used to calculate users' preference vector. Furthermore, each participant's preference vector would also be needed in order to generate subsequent recommendation for both collaborative filtering and content-based recommender systems. In this study, participants were presented a questionnaire containing a list of the 23 genres at the beginning of the experiment to obtain their preference vectors. Participants were instructed to rate each genre according to their preference levels on a 5-point scale, with 1 representing extremely dislike and 5 representing extremely like. The results of the questionnaire formed the preference of each participant, and were used in both the collaborative filtering and content-based recommender systems. Algorithms for both collaborative filtering and content-based were generic algorithms adopted from Wu, Joung and Lee (2011).

In addition, considering the "Anomalies and exceptions" approach in our study, we also introduced poor similarity and randomness in both collaborative filtering and content-based algorithms. The algorithms are classified as follows: collaborative filtering (CF), collaborative filtering with less precision (CFL), collaborative filtering with high rating less precision (CFHL), content-based (CB), content-based with less precision (CBL) and content-based with high rating less precision (CBHL). Collaborative filtering algorithms are discussed first below and then content-based algorithms.

Collaborative filtering (CF) recommended the most popular products that had been purchased by users of the similar interest group, namely recommendations were made solely on the basis of similarities to other users. First, we used Pearson's correlation to measure similarity between participant *p* and each Netflix user *u*. Let *p*.Pref denote the preference vector of participant *p*, and *p*.Pref[i] representing participant *p*'s preference level towards genre *i*. A participant p's interest group was determined as follows:

$$\mathrm{sim}(p,u) = \frac{\sum_j p.\mathrm{Pref}[j] \times u.\mathrm{Pref}[j] - \frac{\sum_j p.\mathrm{Pref}[j] \times \sum_j u.\mathrm{Pref}[j]}{n}}{\sqrt{\sum_j p.\mathrm{Pref}^2[j] - \frac{(\sum_j p.\mathrm{Pref}[j])^2}{n}} \times \sqrt{\sum_j u.\mathrm{Pref}^2[j] - \frac{(\sum_j u.\mathrm{Pref}[j])^2}{n}}}$$

Let u_m be the Netflix user with the highest similarity with the participant *p*. Then, *p*'s interest group was $\{u \in \mathbf{U} \mid \mathrm{sim}(p, u) \geq 0.8 \times \mathrm{sim}(p, u_m)\}$ where **U** is the set of

Netflix users. Next, each movie had a score predicted by Sum of Ratings (Wu, Joung & Lee, 2011) added up all the rating scores for each movie in the interest group. The movies were ranked and sorted in a descending order according to their respective scores, and the top 10 movies were chosen and recommended to the participant.

Further, concerning collaborative filtering with less precision (CFL) and collaborative filtering with high rating less precision (CFHL) approaches, the top movies were still chosen but not the most “top” anymore. Depends on the movies were sorted in a descending order, the top 10% movies were chosen instead and composed a candidate set. Collaborative filtering with less precision (CFL) recommended 10 movies from this candidate set randomly and each movie in the candidate had equal probability to be recommended. In addition, the movies in the interest group which average rating was greater than 4 were selected and sorted by Sum of Ratings in a descending order. Then the top 10% movies where chosen to compose a high rating candidate set in the same way. Identically, Collaborative filtering with high rating less precision (CFHL) recommended 10 movies randomly from the high rating candidate set in an equal probability.

For content-based (CB) recommender which recommended the movie that best matches a user’s preference, similarity was first calculated between the participant p and each movie x:

$$\mathrm{sim}(p, x) = \frac{\sum_j p.\mathrm{Pref}[j] \times x.\mathrm{Attr}[j]}{\sum_j x.\mathrm{Attr}[j]}$$

Let x_m be the movie with the highest similarity with the participant p. Then, p's candidate set of movies was defined as $\{x \in \mathbf{M} \mid \mathrm{sim}(p, x) \geq 0.8 \times \mathrm{sim}(p, x_m)\}$ where $\mathbf{M}$ is the set of movies. After p's candidate set have been formed, the average ratings of each movie in the candidate set rated by every Netflix user were calculated to form a composite score for each movie. Similar to the collaborative filtering (Classic) recommender, the movies were then ranked according to their respective scores, and the top 10 movies were recommended to the participant.

In consider to content-based with less precision (CBL) and content-based with high rating less precision (CBHL), the recommendations would not always be the case of most "top" movies again. For content-based with less precision (CBL), 10 movies from the participant p's candidate set of movies were selected and recommended to the participant randomly in equal probability. For content-based with high rating less precision (CBHL), the participant p's candidate set of movies was defined as $\{x \in \mathbf{M_H} \mid \mathrm{sim}(p, x) \geq 0.8 \times \mathrm{sim}(p, x_m)\}$ where $\mathbf{M_H}$ is the set of movies which average rating was greater than 4. Similarly, 10 random movies were then recommended in like manner.

3.2. Experiment Procedure

Participants were invited to our laboratory to conduct the experiment. They were asked to fill a written consent form and were then instructed of the experiment procedure. The experiment was divided into three stages. At the first stage, the participants were asked to take a pretest to complete a questionnaire to obtain data on their preferences towards movies in order for the recommendation systems to have sufficient predictive abilities and eliminate a possible cold-start problem.

After the participants finished taking the preference questionnaire, they were then randomly assigned to a recommender strategy and were led into the second stage of our experiment. In this stage, participants were presented an online DVD rental website, named “imMovies”, specifically designed for the experiment. Particularly, in order to simulate consumer purchase decision process, we adopted the EKB model, a complete and systematic theory model concerning consumer behavior which was presented by Engel, Kollat and Blackwell in 1968. The center concept of the EKB model means consumer purchase decision processing is also problem-solving processing and includes five stages: motivation and recognition of need, information search, evaluate alternatives, purchase, and result of purchase (Engel et al., 1993). Thus, we used the EKB model as guideline and designed our online DVD rental website. The website provided an advance searching mechanism that helped participants filter the movie catalog and search information easily according to specific keywords and dimensions such as movie title, actor/actress or director, etc. In addition, the website also contained a browsing option that allowed the participants to browse movies and locate the movies either by genre, English title, Mandarin title, release year or a combination of the above attributes.

In the second stage, participants were instructed to rent DVD movies due to their interest from the DVD rental website, and there was no limitation on the number of rented DVD movies. Participants were not given any particular purchase scenario to avoid any possible biases induced by the scenarios. After participants have chosen to either search or browse DVD movies, they were then presented with at most 10 results ranked by average Netflix ratings according to their chosen criteria. In addition to the 10 search/browse results, 10 DVD movies recommended by their respective recommenders were also presented in this page. Participants could freely decide whether to add any number of DVDs from the results list or the recommendations list to their shopping cart, or go back to the homepage to begin a new search/browse activity. Note that the same 10 recommendations were presented to the participant regardless of his or her searching or browsing tasks.

After the participants have decided on final DVDs to rent in their shopping cart, they were then taken to the third stage of the experiment and were asked to complete a questionnaire to obtain their serendipity and interest levels for the final chosen DVD movies. Participants assigned to recommenders except the browsing condition were also asked to report has ever known before, interest, willingness to rent levels and overall satisfaction for the 10 recommended DVD movies presented to them. The measurements for all the dependent variables will be discussed in more detail in Section 3.4.2. Finally, they were asked to report their demographic information. The entire experiment took roughly 30 minutes to complete, and each participant were given NT$100 as reward for their participation.

3.3. Participants

A total of 454 participants were recruited from Taipei, Taiwan to attend the experiment. Of them, 320 (70.5%) were valid samples, with 40 samples in each of the 8 experiment conditions. Among them, there were 159 (49.7%) male and 161 (50.3%) female; 263 (82.2%) were students, and 57 (17.8%) were non-students. 53 (16.6%) were aged 19 years or below, 196 (61.2%) between 20 and 24 years old, and 71 (22.2%) aged 25 years or above. In addition, the average number of DVD movies rented was 4.03.

3.4. Measurements

3.4.1. Independent Variable

In this study, the only one independent variable was recommender strategies. As mentioned in the previous section, both classic recommenders (include CB and CF) and recommenders with less precision (include CBL, CFL, CBHL and CFHL) were adopted for the recommender independent variable. In addition, random recommender which conducted the “Role of chance” approach and no recommender (the browsing condition) were also included.

3.4.2. Dependent Variables

As stated in the literature review, both implicit and explicit measures of serendipity, user satisfaction and willingness to pay were utilized in our study. Therefore, the dependent variables of our study were: number of recommended products put into shopping cart (implicit), number of recommended products purchased (implicit), interest towards recommended products (explicit), willingness to rent recommended products (explicit), serendipity level of the recommended products (implicit), overall satisfaction towards recommended products (explicit), percentage of recommended products in the shopping cart (implicit), percentage of recommended products purchased (implicit), interest level towards final chosen products (explicit) and serendipity level of final chosen products (implicit).

Number of recommended products put into shopping cart refers to the number of recommended DVDs that were chosen as one of the items in the candidate pool, whereas number of recommended products purchased refers to the number of recommended DVDs that were actually purchased by the participant. For interest towards recommended products and willingness to rent recommended products, we adopted self report and single-item measures that specifically asked the user how interested he or she was with each of the 10 recommended DVD movies and how willing he or she was to rent each of the 10 recommended DVD movies separately. For interest towards recommended products, measures were obtained using a rating scale from 1 to 100, the higher represented strong interest towards recommendations. The large scale could measure interest level accurately and facilitate to obtain users' serendipity level. For willingness to rent recommended products, measures were obtained using a semantic scale ranging from 1 to 5, with 1 being "Extremely Unsatisfied" and 5 being "Extreme Satisfied". For overall satisfaction towards recommended products, we also adopted single-item measures that specifically asked the user how satisfied he or she was with the 10 recommended DVD movies. Likewise, measures were obtained using a semantic scale ranging from 1 to 5, with 1 being "Extremely Low" and 5 being "Extremely High". Satisfaction towards each recommended products were not evaluated because it seemed a repeated measure compared to interest towards recommended products. For serendipity level of the recommended products, instead of directly asking the how surprised he or she was with each of the 10 recommended DVD movies, we adopted single-item measures that specifically asked the user has he or she ever known of each of the 10 recommended DVD movies. Measures were distinguished on a nominal scale: has seen before, has known before and hasn't known before. DVD Movies were considered to be serendipity if and only if it belongs to unknown and its interest level greater than 50.

Next, since there was no limitation on the number of final chosen products in our study, percentage of recommended products in the shopping cart and percentage of recommended products purchased were also measured. The former refers to the percentage of products in the shopping cart that were recommended; similarly, the latter refers to the percentage of final chosen products that were recommended. Interest towards final chosen products was measured in identical ways as interest towards recommended products. Self report rating scale measures from 1 to 100 were used to assess the participants' interest level towards each of their own final chosen products. Likewise, serendipity level of final chosen products was measured in the same way as serendipity level of the recommended products. Interest level towards final chosen products and nominal scale measures were also used to assess the participants' serendipity level of each of their final chosen products. In addition, satisfaction towards final chosen products and willingness to rent final chosen products were not considered in our study. Since the user had already decided to pay for their final chosen products, their satisfaction and willingness to rent should be high and there was no necessary to evaluate these two metrics. Please refer to

Appendix A for a detailed list of the measurement items of the dependent variables.

Due to the nature of our hypotheses, different product level serendipity measures were used to analyze the first four hypotheses. For Hypotheses 1 and 2, we would like to compare serendipity level of the recommended products among classic recommenders (include CF and CB), recommenders with less precision (include CBL, CFL, CBHL and CFHL) and random recommender. The main objective of Hypotheses 1 and 2 was to examine the impact of the “Role of chance” approach work on random recommender and the “Anomalies and exceptions” approach work on both collaborative filtering and content-based recommenders. Subsequently, we would like to compare serendipity on consumption behavior between the various recommender strategies mentioned above. Therefore, our product-level serendipity analyses for Hypotheses 3 and 4 were specifically tailored towards the final chosen products.

Hypotheses sets 5 and 6 considers the effect “Role of chance” approach and “Anomalies and exceptions” approach have on product-level interest, user satisfaction and willingness to pay. The metrics analyses for Hypotheses sets 5 and 6 were tested using interest level towards recommended products, overall satisfaction towards recommended products, willingness to rent recommended products, number of recommended products put into shopping cart and number of recommended products purchased respectively. In addition, for Hypotheses set 7 and 8, we were interested in examining the effects of the approaches mentioned above have on product level interest and user consumption behavior. Similar measures with Hypotheses set 5 and 6 were used to test both of these hypotheses set, which were Interest level towards final chosen products, percentage of recommended products in the shopping cart and percentage of recommended products purchased respectively. Please see Table 3-3 for a detailed description of the dependent variables and their corresponding hypotheses. We will refer to our analysis for Hypotheses 1, 2, 5 and 6 as “Recommender-Specific Results” and Hypotheses 3, 4, 7 and 8 as “Shopping Cart-Specific Results” in the next chapter.

Dependent Variable	**Description**	**Hypothesis**
Serendipity level of the recommended products	Interest level and nominal scales for each of the 10 recommended products	Hypothesis 1 Hypothesis 2
Serendipity level of final chosen products	Interest level and nominal scales for each of the final chosen products	Hypothesis 3 Hypothesis 4
Interest level towards recommended products	Single-item 5-point semantic scales for each of the 10 recommended products	Hypothesis 5a Hypothesis 6a
Overall satisfaction towards recommended products	Single-item 5-point semantic scales for the 10 recommended products	Hypothesis 5b Hypothesis 6b
Willingness to rent recommended products	Single-item 5-point semantic scales for each of the 10 recommended products	Hypothesis 5c Hypothesis 6c
Number of recommended products put into shopping cart	Number of recommended products chosen in the candidate pool, ranging from 0 to 10	Hypothesis 5d Hypothesis 6d
Number of recommended products purchased	Number of recommended products rented out of the final chosen products, ranging from 0 to 10	Hypothesis 5d Hypothesis 6d

Interest level towards final chosen products	Single-item 5-point semantic scales for each of the final chosen products	Hypothesis 7a Hypothesis 8a
Percentage of recommended products in the shopping cart	Percentage of products in the shopping cart that are recommended, ranging from 0 to 100	Hypothesis 7b Hypothesis 8b
Percentage of recommended products purchased	Percentage of final chosen products that are recommended, ranging from 0 to 100	Hypothesis 7b Hypothesis 8b

Table 3-3: Description of the Dependent Variables

4. Empirical Results

ANOVA was used to test all our hypotheses. As mentioned in the end of previous chapter, our results were presented in two separate analyses, with Hypotheses 1, 2, 5 and 6 presented in Section 4.1 (Recommender-Specific Results) and Hypotheses 3, 4, 7 and 8 presented in Section 4.2 (Shopping Cart-Specific Results).

Before we present our results, we would like to illustrate several metrics to present our results. First, due to there was no limitation on the number of final chosen products in the shopping cart, dependent variables such as serendipity level of final chosen products, percentage of recommended products in the shopping cart and percentage of recommended products purchased were all proportions data. Thus, we conducted angular or arcsine transformation to transform the above dependent variables. Let x be the original observation data and y be the observation data after transformation. Then, y was equal to $\sin^{-1}\sqrt{x}$ (Scale of y from 0 to 90).

Secondly, the definition of serendipity in our study was movies which were unknown to the user and interest levels were greater than 50. As mentioned before, random recommender was set as a baseline to ensure all the other recommenders were actually effective. The average interest level towards recommended products of random recommender was 46.21, thus we chose 50 as the threshold to determine serendipity.

In addition, since our sample was highly consisted of students, two-sample Kolmogorov-Smirnov test for student and non-student groups were assessed on all our dependent variables. According to Table 4-1, there were no significant differences (p's >= .088) in sample distribution between students and non-students for all dependent variables.

Dependent Variable	**KS**	**D**	**KSa**	**Pr > KSa**
Serendipity level of the recommended products	.009	.022	.451	.987
Serendipity level of final chosen products	.034	.088	.605	.857
Interest level towards recommended products	.022	.057	1.180	.124
Overall satisfaction towards recommended products	.026	.066	.437	.991
Willingness to rent recommended products	.017	.042	.880	.420
Number of recommended products put into shopping cart	.040	.102	.674	.754
Number of recommended products purchased	.037	.095	.626	.828
Interest level towards final chosen products	.035	.090	1.249	.088
Percentage of recommended products in the shopping cart	.046	.116	.765	.602
Percentage of recommended products purchased	.062	.157	1.033	.236

Table 4-1: Two-sample K-S tests between students and non-students

4.1. Recommender-Specific Results

In our analyses for recommender-specific results, we will submit 280 samples under 7 possible experiment conditions (random, content-based, collaborative filtering, content-based with less precision, collaborative filtering with less precision, content-based with high rating less precision, collaborative filtering with high rating less precision) to ANOVA. For Hypotheses 5a, 5c, 6a and 6c, subject random factor was added into the model to control for variances between the subjects. Due to the resulting mixed-models for these hypotheses, quasi F’ values were reported for their respective assessments. Standard F value was used to assess Hypothesis 1, 2, 5b, 5d, 6b and 6d.

4.1.1. Serendipity level of the recommended products

CB	CF	RAN	CBL	CFL	CBHL	CFHL
1.30 (1.42)	1.08 (1.44)	3.03 (2.89)	3.10 (2.53)	2.35 (2.01)	3.30 (2.54)	2.63 (2.41)

Table 4-2: Mean and SD for Serendipity level of the recommended products

Recommender	N	Mean	SD	*t*(78)	p-value
CBL	40	3.10	2.53	0.12	.45
CFL	40	2.35	2.01	1.21	.11
CBHL	40	3.30	2.54	0.45	.33
CFHL	40	2.63	2.41	0.67	.25

Table 4-3: Priori pair wise T-test results of Serendipity level of the recommended products between random recommender (n = 40, M = 3.03, SD = 2.89) and recommenders with less precision

Firstly, for the serendipity level of the recommended products dependent variable, differences between each recommender conditions were statistically significant ($F(6, 273) = 6.25$, $MSe = 5.021$, $p < .0001$). Please see Table 4-2 for means and standard deviations for the recommender effect on serendipity level of the recommended product. A priori pair wise T-test comparisons revealed that the Random (n = 40, M = 3.03, SD = 2.89) group resulted in significantly higher serendipity level of their recommended products than both the collaborative-filtering (n = 40, M = 1.08, SD = 1.44, $t(78) = 3.82$, $p < .001$) and content-based (n = 40, M = 1.30, SD = 1.42, $t(78)$ =

3.39, $p < .001$) groups. Results showed strong support for Hypothesis 1.

Next, we focused on the difference of serendipity level between recommenders with less precision and classic recommenders. A priori pair wise T-test comparisons revealed that both the collaborative-filtering with less precision ($n = 40$, $M = 2.35$, $SD = 2.01$, $t(78) = 3.27$, $p < .001$) and collaborative-filtering with high rating less precision ($n = 40$, $M = 2.63$, $SD = 2.41$, $t(78) = 3.49$, $p < .001$) groups resulted in significantly higher serendipity level of their recommended products than the collaborative-filtering group ($n = 40$, $M = 1.08$, $SD = 1.44$). In addition, participants assigned to the collaborative-filtering with less precision and collaborative-filtering with high rating less precision groups showed no significant difference ($t(78) = 0.55$, $p = .58$) in their serendipity level of the recommended products. Identically, both the content-based with less precision ($n = 40$, $M = 3.10$, $SD = 2.53$, $t(78) = 3.93$, $p < .0001$) and content-based with high rating less precision ($n = 40$, $M = 3.30$, $SD = 2.54$, $t(78) = 4.34$, $p < .0001$) groups resulted in significantly higher serendipity level of their recommended products than the content-based group ($n = 40$, $M = 1.30$, $SD = 1.42$). In addition, participants assigned to the content-based with less precision and content-based with high rating less precision groups showed no significant difference ($t(78) = 0.35$, $p = .73$) in their serendipity level of the recommended products. Thus, Hypothesis 2 was also strongly supported.

Furthermore, participants assigned to the random recommender and the recommenders with less precision groups showed no significant difference in their serendipity level of the recommended products. Please see Table 4-3 for priori pair wise T-test results of serendipity level of the recommended products between random recommender and recommenders with less precision.

4.1.2. Interest level towards recommended products

CB	CF	RAN	CBL	CFL	CBHL	CFHL
64.15 (23.33)	65.85 (22.99)	46.21 (24.94)	52.60 (25.08)	64.07 (22.97)	57.92 (23.16)	64.93 (22.54)

Table 4-4: Mean and SD for Interest level towards recommended products

Recommender	N	Mean	SD	*t*(78)	p-value
CBL	400	52.60	25.08	3.62	< .001***
CFL	400	64.07	22.97	10.54	< .001***
CBHL	400	57.92	23.16	6.88	< .001***
CFHL	400	64.93	22.54	11.14	< .001***

Table 4-5: Priori pair wise T-test results of serendipity level of the recommended products between random recommender (n = 400, M = 46.21, SD = 24.94) and recommenders with less precision

Secondly, for the interest level towards recommended products dependent variable, differences between the 7 recommender conditions were also statistically significant ($F'(6, 273) = 9.51$, $MSe = 2376.263$, $p < .0001$). Please see Table 4-4 for means and standard deviations for the recommender effect on interest level towards recommended product. A priori pair wise T-test comparisons revealed that both the collaborative-filtering (n = 400, M = 65.85, SD = 22.99, $t(798) = 11.58$, $p < .0001$) and content-based (n = 400, M = 64.15, SD = 23.33, $t(798) = 10.51$, $p < .0001$) groups resulted in significantly higher interest level towards their recommended products than the Random (n = 400, M = 46.21, SD = 24.94) group. Results showed strong support for Hypothesis 5a.

Next, we focused on the difference of interest level between recommenders with less precision and classic recommenders. A priori pair wise T-test comparisons revealed that both the collaborative-filtering with less precision (n = 400, M = 64.07, SD = 22.97, $t(798)$ = 1.10, p = .14) and collaborative-filtering with high rating less precision (n = 400, M = 64.93, SD = 22.54, $t(798)$ = 0.58, p = .28) groups resulted in insignificant difference of interest level than the collaborative-filtering group (n = 400, M = 65.85, SD = 22.99). In addition, participants assigned to the collaborative-filtering with less precision and collaborative-filtering with high rating less precision groups also showed no significant difference ($t(798)$ = 0.53, p = .60) in their interest level towards recommended products. In contrast, surprisingly, both the content-based with less precision (n = 400, M = 52.60, SD = 25.08, $t(798)$ = 6.74, p < .0001) and content-based with high rating less precision (n = 400, M = 57.92, SD = 23.16, $t(798)$ = 3.79, p < .0001) groups resulted in significantly lower interest level towards their recommended products than the content-based group (n = 400, M = 64.15, SD = 23.33). In addition, participants assigned to the content-based with less precision and content-based with high rating less precision groups showed significant difference ($t(798) = 3.11$, $p < .01$) in their interest level towards recommended products. Thus, Hypothesis 6a was not supported.

Furthermore, participants assigned to the recommenders with less precision groups showed significantly higher interest level towards recommended products than the random recommender group. Please see Table 4-5 for priori pair wise T-test results of interest level towards recommended products between random recommender and recommenders with less precision.

4.1.3. Overall satisfaction towards recommended products

CB	CF	RAN	CBL	CFL	CBHL	CFHL
3.30 (0.65)	3.53 (0.85)	2.38 (0.87)	2.73 (0.85)	3.18 (0.78)	2.90 (0.84)	3.25 (0.84)

Table 4-6: Mean and SD for Overall satisfaction towards recommended products

Recommender	N	Mean	SD	*t*(78)	p-value
CBL	40	2.73	0.85	1.83	< .05*
CFL	40	3.18	0.78	4.33	<.001***
CBHL	40	2.90	0.84	2.75	<.01**
CFHL	40	3.25	0.84	4.58	<.001***

Table 4-7: Priori pair wise T-test results of Overall satisfaction towards recommended products between random recommender ($n = 40$, $M = 2.38$, $SD = 0.87$) and recommenders with less precision

Thirdly, for the overall satisfaction towards recommended products dependent variable, differences between the 7 recommender conditions were also statistically significant ($F(6, 273) = 9.33$, $MSe = 0.662$, $p < .0001$). Please see Table 4-6 for means and standard deviations for the recommender effect on overall satisfaction towards recommended product. A priori pair wise T-test comparisons revealed that both the collaborative-filtering (n = 40, M = 3.53, SD = 0.85, $t(78)$ = 6.00, p < .0001) and content-based (n = 40, M = 3.30, SD = 0.65, $t(78)$ = 5.40, p < .0001) groups resulted in significantly higher overall satisfaction towards their recommended products than the Random (n = 40, M = 2.38, SD = 0.87) group. Results showed strong support for Hypothesis 5b.

Next, we focused on the difference of overall satisfaction between recommenders with less precision and classic recommenders. A priori pair wise T-test comparisons revealed that both the collaborative-filtering with less precision (n = 40, M = 3.18, SD = 0.78, $t(78)$ = 1.92, p < .05) and collaborative-filtering with high rating less precision (n = 40, M = 3.25, SD = 0.84, $t(78)$ = 1.46, p = .07) groups resulted in significantly lower and moderately significant lower overall satisfaction towards recommended products than the collaborative-filtering group (n = 40, M = 3.53, SD = 0.85) separately. In addition, participants assigned to the collaborative-filtering with less precision and collaborative-filtering with high rating less precision groups also showed no significant difference ($t(78)$ = 0.41, p = .68) in their overall satisfaction towards recommended products. Identically, both the content-based with less precision (n = 40, M = 2.73, SD = 0.85, $t(78)$ = 3.41, p < .001) and content-based with high rating less precision (n = 40, M = 2.90, SD = 0.84, $t(78)$ = 2.38, p < .01) groups resulted in significantly lower overall satisfaction towards recommended products than the content-based group (n = 40, M = 3.30, SD = 0.65). In addition, participants assigned to the content-based with less precision and content-based with high rating less precision groups showed no significant difference ($t(78) = 0.93$, $p = .36$) in their overall satisfaction towards recommended products. Thus, Hypothesis 6b was not supported.

Furthermore, participants assigned to the recommenders with less precision groups showed significantly higher overall satisfaction towards recommended products than the random recommender group. Please see Table 4-7 for priori pair wise T-test results of overall satisfaction towards recommended products between random recommender and recommenders with less precision.

4.1.4. Willingness to rent recommended products

CB	CF	RAN	CBL	CFL	CBHL	CFHL
3.00 (1.15)	3.06 (1.09)	2.38 (1.04)	2.63 (1.10)	2.99 (1.05)	2.77 (1.05)	3.03 (1.07)

Table 4-8: Mean and SD for Willingness to rent recommended products

Recommender	N	Mean	SD	$t(78)$	p-value
CBL	400	2.63	1.10	3.28	< .001***
CFL	400	2.99	1.05	8.19	< .001***
CBHL	400	2.77	1.05	5.22	< .001***
CFHL	400	3.03	1.07	8.62	< .001***

Table 4-9: Priori pair wise T-test results of Willingness to rent recommended products between random recommender (n = 400, M = 2.38, SD = 1.04) and recommenders with less precision

Fourthly, for the willingness to rent recommended products dependent variable, differences between the 7 recommender conditions were also statistically significant ($F'(6, 273) = 8.19$, $MSe = 3.146$, $p < .0001$). Please see Table 4-8 for means and standard deviations for the recommender effect on willingness to rent recommended product. A priori pair wise T-test comparisons revealed that both the collaborative-filtering ($n = 400$, $M = 3.06$, $SD = 1.09$, $t(798) = 8.93$, $p < .0001$) and content-based ($n = 400$, $M = 3.00$, $SD = 1.15$, $t(798) = 7.97$, $p < .0001$) groups resulted in significantly higher willingness to rent their recommended products than the Random ($n = 400$, $M = 2.38$, $SD = 1.04$) group. Results showed strong support for Hypothesis 5c.

Next, we focused on the difference of willingness to rent between recommenders with less precision and classic recommenders. A priori pair wise T-test comparisons revealed that both the collaborative-filtering with less precision ($n = 400$, $M = 2.99$, $SD = 1.05$, $t(798) = 0.89$, $p = .19$) and collaborative-filtering with high rating less precision ($n = 400$, $M = 3.03$, $SD = 1.07$, $t(798) = 0.39$, $p = .35$) groups resulted in insignificant difference of interest level than the collaborative-filtering group ($n = 400$, $M = 3.06$, $SD = 1.09$). In addition, participants assigned to the collaborative-filtering with less precision and collaborative-filtering with high rating less precision groups also showed no significant difference ($t(798) = 0.50$, $p = .62$) in their willingness to rent recommended products. In contrast, surprisingly, both the content-based with less precision (n = 400, M = 2.63, SD = 1.10, $t(798)$ = 4.65, $p < .0001$) and content-based with high rating less precision (n = 400, M = 2.77, SD = 1.05, $t(798)$ = 2.99, $p < .005$) groups resulted in significantly lower willingness to rent their recommended products than the content-based group (n = 400, M = 3.00, SD = 1.15). In addition, participants assigned to the content-based with less precision and content-based with high rating less precision groups showed moderately significant difference ($t(798)$ = 1.81, p = .07) in their willingness to rent recommended products. Thus, Hypothesis 6c was not supported.

Furthermore, participants assigned to the recommenders with less precision groups showed significantly higher willingness to rent recommended products than the random recommender group. Please see Table 4-9 for priori pair wise T-test results of willingness to rent recommended products between random recommender and recommenders with less precision.

4.1.5. Number of recommended products accepted

Number of recommended products put into shopping cart						
CB	**CF**	**RAN**	**CBL**	**CFL**	**CBHL**	**CFHL**
1.53 (1.69)	1.90 (1.66)	0.78 (0.89)	0.83 (1.06)	1.38 (1.08)	1.35 (1.27)	1.35 (1.14)
Number of recommended products purchased						
CB	**CF**	**RAN**	**CBL**	**CFL**	**CBHL**	**CFHL**
1.45 (1.63)	1.58 (1.20)	0.60 (0.84)	0.73 (0.88)	1.18 (1.08)	1.20 (1.24)	1.25 (1.13)

Table 4-10: Mean and SD for Number of recommended products accepted

Number of recommended products put into shopping cart					
Recommender	**N**	**Mean**	**SD**	***t*(78)**	**p-value**
CBL	40	0.83	1.06	0.23	.41
CFL	40	1.38	1.08	2.71	<.01**
CBHL	40	1.35	1.27	2.34	<.05*
CFHL	40	1.35	1.14	2.51	<.01**

Table 4-11: Priori pair wise T-test results of Number of recommended products put into shopping cart between random recommender (n = 40, M = 0.78, SD = 0.89) and recommenders with less precision

Number of recommended products purchased					
Recommender	**N**	**Mean**	**SD**	***t*(78)**	**p-value**
CBL	40	0.73	0.88	0.65	.26
CFL	40	1.18	1.08	2.65	<.01**
CBHL	40	1.20	1.24	2.53	<.01**
CFHL	40	1.25	1.13	2.92	<.01**

Table 4-12: Priori pair wise T-test results of Number of recommended products purchased between random recommender (n = 40, M = 0.60, SD = 0.84) and recommenders with less precision

Finally, for Hypothesis 5d and Hypothesis 6d, we would like to compare the number of recommended products accepted by user between various recommender conditions. We used two dependent variables: number of recommended products put into shopping cart and number of recommended products purchased, to evaluate our hypotheses. For the number of recommended products put into shopping cart dependent variable, differences between the 7 recommender conditions were statistically significant (F(6, 273) = 3.70, MSe = 1.663, p < .005); for the number of recommended products purchased dependent variable, between the 7 recommender conditions were also statistically significant (F(6, 273) = 3.74, MSe = 1.366, p < .005). Please see Table 4-10 for means and standard deviations for the recommender effect on number of recommended products put into shopping cart and number of recommended product purchased. For number of recommended products put into shopping cart, a priori pair wise T-test comparisons revealed that both the collaborative-filtering (n = 40, M = 1.90, SD = 1.66, $t(78)$ = 3.77, p < .0005) and content-based (n = 40, M = 1.53, SD = 1.69, $t(78)$ = 2.48, p < .01) groups resulted in significantly higher number of recommended products put into shopping cart than the Random (n = 40, M = 0.78, SD = 0.89) group. For number of recommended products purchased, a priori pair wise T-test comparisons revealed that both the collaborative-filtering (n = 40, M = 1.58, SD = 1.20, $t(78)$ = 4.22, p < .0001) and content-based (n = 40, M = 1.45, SD = 1.63, $t(78)$ = 2.93, p < .005) groups resulted in significantly higher number of recommended products purchased than the Random (n = 40, M = 0.60, SD = 0.84) group. Results showed strong support for Hypothesis 5d.

Next, we focused on the difference of number of recommended products accepted between recommenders with less precision and classic recommenders. For number of recommended products put into shopping cart, a priori pair wise T-test comparisons revealed that both the collaborative-filtering with less precision (n = 40, M = 1.38, SD = 1.08, $t(78)$ = 1.68, p < .05) and collaborative-filtering with high rating less precision (n = 40, M = 1.35, SD = 1.14, $t(78)$ = 1.72, p < .05) groups resulted in significant lower number of recommended products put into shopping cart than the collaborative-filtering group (*n* = 40, *M* = 1.90, *SD* = 1.66). In addition, participants assigned to the collaborative-filtering with less precision and collaborative-filtering with high rating less precision groups also showed no significant difference ($t(78)$ = 0.10, p = .92) in their number of recommended products put into shopping cart. Identically, the content-based with less precision (*n* = 40, *M* = 0.83, *SD* = 1.06, $t(78)$ = 2.22, $p < .05$) and content-based with high rating less precision (*n* = 40, *M* = 1.35, *SD* = 1.27, $t(78)$ = 0.52, p = .30) groups resulted in significantly lower and no significant difference of number of their recommended products put into shopping cart separately compared with the content-based group (*n* = 40, *M* = 1.53, *SD* = 1.69). In addition, participants assigned to the content-based with less precision and content-based with high rating less precision groups showed significant difference ($t(78)$ = 2.01, $p < .05$) in their number of recommended products put into shopping cart.

For number of recommended products purchased, a priori pair wise T-test comparisons revealed that both the collaborative-filtering with less precision (*n* = 40, *M* = 1.18, *SD* = 1.08, $t(78)$ = 1.57, p = .06) and collaborative-filtering with high rating less precision (*n* = 40, *M* = 1.25, *SD* = 1.13, $t(78)$ = 1.25, p = .11) groups resulted in moderately significantly lower and no significant difference of number of recommended products purchased separately compared to the collaborative-filtering group (*n* = 40, *M* = 1.58, *SD* = 1.20). In addition, participants assigned to the collaborative-filtering with less precision and collaborative-filtering with high rating less precision groups also showed no significant difference ($t(78) = 0.30$, $p = .76$) in their number of recommended products purchased. Identically, the content-based with less precision ($n = 40$, $M = 0.73$, $SD = 0.88$, $t(78) = 2.47$, $p < .01$) and content-based with high rating less precision ($n = 40$, $M = 1.20$, $SD = 1.24$, $t(78) = 0.77$, $p = .22$) groups resulted in significantly lower and no significantly difference of number of their recommended products purchased separately with the content-based group ($n = 40$, $M = 1.45$, $SD = 1.63$). In addition, participants assigned to the content-based with less precision and content-based with high rating less precision groups showed moderately significant difference ($t(78) = 1.97$, $p = .05$) in their number of recommended products purchased. Thus, Hypothesis 6d was not supported.

Furthermore, participants assigned to the recommenders with less precision groups showed significantly higher number of recommended products accepted than the random recommender group (except the CBL group). Please see Table 4-11 and Table 4-12 for priori pair wise T-test results of number of recommended products accepted between random recommender and recommenders with less precision.

In general, results indicated support for Hypotheses 1, 2 and Hypotheses set 5, but not for Hypotheses set 6. In particular, participants expressed significantly higher serendipity level of the recommended products for both the random recommender and the recommenders with less precision groups compared to the classic recommenders group, showing support for Hypotheses 1 and 2. In addition, there were no differences in serendipity level of the recommended products between the random recommender and the recommenders with less precision groups.

Further, participants expressed significantly lower interest level towards recommended products, overall satisfaction towards recommended products, willingness to rent recommended products and number of recommended products accepted for the random recommender group compared to the classic recommenders group, showing support for Hypotheses set 5. We also evaluated the differences of the above metrics between the random recommender group and the recommenders with less precision group. Identically, results indicated that participants much preferred the recommendations of the recommenders with less precision. In addition, participants expressed significantly higher interest level towards recommended products, willingness to rent recommended products and number of recommended products accepted for the content-based recommender with high rating less precision group compared to the content-based recommender with less precision group, showing "high rating" take effect on content-based recommender particularly.

In contrast, participants expressed either significantly lower or no differences interest level towards recommended products, overall satisfaction towards recommended products, willingness to rent recommended products and number of recommended products accepted for the recommenders with less precision group and the recommenders with high rating less precision group compared to the classic recommenders group. This result indicated that the “Anomalies and exceptions” approach didn’t make a positive impact on the quality criteria for classic recommenders. However, for the above metrics, the result also indicated that the content-based recommenders with less precision and high rating randomness compared to the classic content-based recommender suffered from much more significant differences than the collaborative-filtering recommenders with less precision and high rating randomness compared to the classic collaborative-filtering recommender; namely, the quality of recommendations of collaborative-filtering recommenders were less affected by the “Anomalies and exceptions” approach.

4.2. Shopping Cart-Specific Results

In our analyses for shopping cart-specific results, for Hypothesis 3, 4, 7a and 8a, all 320 samples under 8 possible experiment conditions (random, content-based, collaborative filtering, content-based with less precision, collaborative filtering with less precision, content-based with high rating less precision, collaborative filtering with high rating less precision, no recommender) were submitted into ANOVA. For Hypothesis 7b and 8b, the “no recommender” condition was excluded and only 280 samples were submitted. In addition, subject random factor was added into the model to control for variances between the subjects for Hypotheses 7a and 8a. Due to the resulting mixed-models for these hypotheses, quasi F' values were reported for their respective assessments. Standard F value was used to assess Hypothesis 3, 4, 7b and 8b.

4.2.1. Serendipity level of final chosen products

CB	CF	RAN	CBL	CFL	CBHL	CFHL	NO
19.77 (25.10)	19.91 (22.22)	24.55 (29.10)	24.61 (29.40)	19.99 (23.22)	26.65 (24.80)	24.92 (22.98)	15.66 (23.06)

Table 4-13: Mean and SD for Serendipity level of final chosen products

Firstly, for the serendipity level of the final chosen products dependent variable, differences between each recommender conditions were no significant difference ($F(7, 312) = 0.88$, $MSe = 631.138$, $p = .52$). Please see Table 4-13 for means and standard deviations for the recommender effect on serendipity level of the final chosen product. Thus, Hypothesis 3 and Hypothesis 4 were not supported.

4.2.2. Interest level towards final chosen products

CB	CF	RAN	CBL	CFL	CBHL	CFHL	NO
84.13 (12.42)	81.54 (12.77)	79.68 (11.54)	82.57 (10.70)	81.28 (11.69)	83.01 (8.34)	82.04 (11.54)	82.79 (11.59)

Table 4-14: Mean and SD for Interest level towards final chosen products

Next, for the interest level towards final chosen products dependent variable, difference between each recommender conditions also showed no significant difference ($F'(7, 360.92) = 0.79$, $MSe = 242.779$, $p = .59$). Please see Table 4-14 for means and standard deviations for the recommender effect on interest level of the final chosen product. Thus, Hypothesis 7a and Hypothesis 8a were not supported.

4.2.3. Percentage of recommended products accepted

Percentage of recommended products in the shopping cart						
CB	**CF**	**RAN**	**CBL**	**CFL**	**CBHL**	**CFHL**
33.08 (32.29)	39.25 (25.40)	16.19 (17.37)	22.18 (28.42)	30.48 (21.41)	31.23 (26.82)	32.15 (27.16)
Percentage of recommended products purchased						
CB	**CF**	**RAN**	**CBL**	**CFL**	**CBHL**	**CFHL**
34.27 (33.86)	39.16 (26.83)	14.82 (18.65)	22.53 (28.44)	28.29 (23.15)	31.06 (26.93)	32.16 (27.89)

Table 4-15: Mean and SD for Percentage of recommended products accepted

Percentage of recommended products in the shopping cart					
Recommender	**N**	**Mean**	**SD**	***t*(78)**	**p-value**
CBL	40	22.18	28.42	1.14	.13
CFL	40	30.48	21.41	3.28	<.001***
CBHL	40	31.23	26.82	2.98	<.01**
CFHL	40	32.15	27.16	3.13	<.01**

Table 4-16: Priori pair wise T-test results of Percentage of recommended products in the shopping cart between random recommender ($n = 40$, $M = 16.19$, $SD = 17.37$) and recommenders with less precision

Percentage of recommended products purchased					
Recommender	**N**	**Mean**	**SD**	***t*(78)**	**p-value**
CBL	40	22.53	28.44	1.43	.08
CFL	40	28.29	23.15	2.87	<.01**
CBHL	40	31.06	26.93	3.14	<.01**
CFHL	40	32.16	27.89	3.27	<.001***

Table 4-17: Priori pair wise T-test results of Percentage of recommended products purchased between random recommender ($n = 40$, $M = 14.82$, $SD = 18.65$) and recommenders with less precision

Finally, for Hypothesis 7b and Hypothesis 8b, we would like to further examine the results whether the products that the participant has chosen were of recommended or not. We used two dependent variables: percentage of recommended products in the shopping cart and percentage of recommended products purchased, to evaluate our hypotheses. For the percentage of recommended products in the shopping cart dependent variable, differences between the 7 recommender conditions were statistically significant ($F(6, 273) = 3.46$, $MSe = 673.214$, $p < .005$); for the percentage of recommended products purchased dependent variable, between the 7 recommender conditions were also statistically significant ($F(6, 273) = 3.58$, $MSe = 723.118$, $p < .005$). Please see Table 4-15 for means and standard deviations for the recommender effect on percentage of recommended products in the shopping cart and percentage of recommended products purchased. For percentage of recommended products in the shopping cart, a priori pair wise T-test comparisons revealed that both the collaborative-filtering ($n = 40$, $M = 39.25$, $SD = 25.40$, $t(78) = 4.74$, $p < .0001$) and content-based ($n = 40$, $M = 33.08$, $SD = 32.29$, $t(78) = 2.91$, $p < .005$) groups resulted in significantly higher number of recommended products put into shopping cart than the Random ($n = 40$, $M = 16.19$, $SD = 17.37$) group. For percentage of recommended products purchased, a priori pair wise T-test comparisons revealed that both the collaborative-filtering ($n = 40$, $M = 39.16$, $SD = 26.83$, $t(78) = 4.71$, $p < .0001$) and content-based ($n = 40$, $M = 34.27$, $SD = 33.86$, $t(78) = 3.18$, $p < .005$)

groups resulted in significantly higher number of recommended products purchased than the Random (n = 40, M = 14.82, SD = 18.65) group. Results showed strong support for Hypothesis 7b.

Next, we focused on the difference of percentage of recommended products accepted between recommenders with less precision and classic recommenders. For percentage of recommended products in the shopping cart, a priori pair wise T-test comparisons revealed that both the collaborative-filtering with less precision (n = 40, M = 30.48, SD = 21.41, $t(78)$ = 1.67, $p < .05$) and collaborative-filtering with high rating less precision (n = 40, M = 32.15, SD = 27.16, $t(78)$ = 1.21, p = .12) groups resulted in significant lower and no significant difference of their percentage of recommended products in the shopping cart separately compared with the collaborative-filtering group (n = 40, M = 39.25, SD = 25.40). In addition, participants assigned to the collaborative-filtering with less precision and collaborative-filtering with high rating less precision groups showed no significant difference ($t(78)$ = 0.31, p = .76) in their percentage of recommended products in the shopping cart. Identically, the content-based with less precision (n = 40, M = 22.18, SD = 28.42, $t(78)$ = 1.60, p = .06) and content-based with high rating less precision (n = 40, M = 31.23, SD = 26.82, $t(78)$ = 0.28, p = .39) groups resulted in moderately significantly lower and no significant difference of number of their recommended products put into shopping cart separately compared with the content-based group (n = 40, M = 33.08, SD = 32.29). In addition, participants assigned to the content-based with less precision and content-based with high rating less precision groups also showed no significant difference ($t(78) = 1.46$, $p = .15$) in their percentage of recommended products in the shopping cart.

For percentage of recommended products purchased, a priori pair wise T-test comparisons revealed that both the collaborative-filtering with less precision (n = 40, M = 28.29, SD = 23.15, $t(78) = 1.94$, $p < .05$) and collaborative-filtering with high rating less precision (n = 40, M = 32.16, SD = 27.89, $t(78) = 1.14$, $p = .13$) groups resulted in significantly lower and no significant difference of their percentage of recommended products purchased separately with the collaborative-filtering group (n = 40, M = 39.16, SD = 26.83). In addition, participants assigned to the collaborative-filtering with less precision and collaborative-filtering with high rating less precision groups showed no significant difference ($t(78) = 0.68$, $p = .50$) in their percentage of recommended products purchased. Identically, the content-based with less precision (n = 40, M = 22.53, SD = 28.44, $t(78) = 1.68$, $p < .05$) and content-based with high rating less precision (n = 40, M = 31.06, SD = 26.93, $t(78)$ = 0.47, $p = .32$) groups resulted in significantly lower and no significantly difference of percentage of recommended products purchased separately compared with the content-based group (n = 40, M = 34.27, SD = 33.86). In addition, participants assigned to the content-based with less precision and content-based with high rating less precision groups also showed no significant difference ($t(78) = 1.38$, $p = .17$) in percentage of recommended products purchased. Thus, Hypothesis 8b was not supported.

Furthermore, participants assigned to the recommenders with less precision groups showed significantly higher percentage of recommended products accepted than the random recommender group (except the CBL group). Please see Table 4-16 and Table 4-17 for priori pair wise T-test results of percentage of recommended products accepted between random recommender and recommenders with less precision.

Overall, results indicated strong support for Hypothesis 7b, but not for Hypotheses 3, 4, 7a and Hypotheses set 8. Participants expressed significantly higher percentage of recommended products accepted for both the classic recommenders and the recommenders with less precision groups compared to the random recommender group, showing support for Hypothesis 7b. In contrast, for participants assigned to the recommenders with less precision or recommenders with high rating less precision, results of percentage of recommended products accepted dependent variable were not consistent with our hypothesis 8b. As the above mentioned, participants assigned to the recommenders with less precision group and the classic recommenders group purchased more recommended movies then the random recommender group, but there was no difference between the recommenders with less precision group and the classic recommenders group.

Further, surprisingly, results of ANOVA on serendipity level of final chosen products and interest level towards final chosen products were all not significant. Participants under the 7 recommender conditions showed no difference on these two dependent variables, indicated that Hypotheses 3, 4, 7a, 8a were not supported. For serendipity level of final chosen products dependent variable, the random recommender group and the recommenders with less precision group garnered higher serendipity level than the classic recommender systems condition, although their differences were insignificant. In the other hand, participants expressed lower interest level towards recommended products of the random recommender group, but the differences were still insignificant. The results pointed out that a movie was serendipitous or not might be not the main concern when participants put in shopping cart. Further, results also showed that for those movies which put in the shopping cart, participants were all interested no matter which recommender condition was.

5. Conclusion

5.1. Discussions

Measuring the impact recommenders have on the success of E-commerce has been an important issue. Previous researches have mostly focused on the efficiency and accuracy of the recommenders, as accuracy has been assumed to lead to increased user satisfaction and purchase intention. However, researches have shown that high recommendation accuracy does not always lead to optimal user satisfaction (McNee, 2002; Ziegler, 2005), and it have been suggested that the impact of recommenders should also be evaluated at a user-level by directly measuring what the user feels after being helped by recommendations (Herlocker et. al., 2004).

Our study focused on how to trigger serendipity of recommendations by adopting the "Role of chance" approach and the "Anomalies and exceptions" approach which introduced by Toms (2000). Next, we directly examined the impact of different approaches of recommender systems in the consumer perspective by measuring several important consumer outcome variables: interest, satisfaction and willingness to pay.

Results indicate that the recommenders which adopted the "Role of chance" approach and the "Anomalies and exceptions" approach increased serendipity level of the recommended products, and there was no difference on serendipity level between these two approaches. This implies that in the E-commerce context, consumers being aided by recommenders with the above two approaches will discover something new, interesting and unexpected products compared to ones that were aided by classic recommenders. However, although random recommender which adopted the “Role of chance” approach increased serendipity level of recommended products, it did not consider the relevance of user preference and only rely on “blind luck”. As a result, the recommended products of random recommender were unfavorable. In contrast, recommenders which adopted the “Anomalies and exceptions” approach provided not only new information to the user but also relevant recommendations and take account of quality.

Next, in the “Anomalies and exceptions” approach, we implemented via recommenders with less precision and recommenders with high rating less precision. The basic difference was the candidate pool. The latter composed its candidate set only for products which average rating was greater than 4. The distinction seemed to be large, whereas recommenders with high rating less precision recommended highly qualified products. Surprisingly, the effect differed between content-based recommenders and collaborative-filtering recommenders. For content-based recommenders with high rating less precision, participants expressed higher interest level, willingness to pay and accepted more recommended products than content-based recommenders with less precision at the recommender-specific stage. In contrast, for collaborative-filtering recommenders, there was no significant difference between less precision and high rating less precision on these metrics and other metrics. Furthermore, for content-based recommenders implemented the “Anomalies and exceptions” approach, participants also expressed significantly lower interest level, willingness to pay and satisfaction toward recommended products than classic content-based recommenders at the recommender-specific stage. Oppositely, for collaborative-filtering recommenders, the only difference between the classic ones and the ones adopted the “Anomalies and exceptions” approach was satisfaction. This implies that the collaborative-filtering recommenders were much unaffected by the drawback of the “Anomalies and exceptions” approach which in order to introduce less well-known and obscure products because of the collaborative-filtering recommenders tend to recommend popular products substantially. For content-based recommenders which adopted the “Anomalies and exceptions” approach, the serendipity level towards recommended products increased indeed, but several metrics for evaluating the quality of recommendations were suffered from dramatic declines.

Moreover, at the shopping cart-specific stage, results still followed the same essential pattern but the relative differences were less. For example, there was no difference in terms of interest level among recommenders with less precision, recommenders with high rating less precision and classic recommenders. Interestingly, the effects on serendipity level of recommended products of the “Role of chance” approach and the “Anomalies and exceptions” approach were unobvious at the shopping cart-specific stage. The results showed the means of serendipity level of the above two approaches were greater than the means of serendipity level of classic recommenders actually, but the differences were not significant. The result implies that participant might not purchase more unexpected products which they were interested.

The two approaches focus on improving serendipity introduced in this study both played important roles in affecting the impact of recommenders. Both of them triggered serendipities to happen but had negative impact on all of the other metrics which traditionally used to measure the quality of recommender systems at the recommender-specific stage. The phenomenon was not consistent with our hypotheses, since serendipitous products could provide new information to the user and also arouse user's interest. We thought there might be a trade-off relationship between serendipity and other metrics used to evaluate recommenders. In this study, we broaden the coverage of recommended products though random and less precision strategies. Truly, the two strategies triggered serendipity by providing new information to the user, but the overall quality of recommendations was sacrificed subsequently.

To conclude, this study contributes to the ongoing research on serendipity of recommendations by providing empirical evidence that the introduction of the “Role of chance” and the “Anomalies and exceptions” approaches increases serendipity level. In addition, this study discovered that the “Anomalies and exceptions” approach is a much suitable method to trigger serendipitous encounters, whereas the “Role of chance” approach simply relies on blind luck. Moreover, for the “Anomalies and exceptions” approach, recommender with high rating less precision had better performance on most of all traditional metrics which used to evaluate recommender systems compared to recommenders with less precision. Finally, results of the study suggest that collaborative-filtering recommender systems are much more preferable to implement the “Anomalies and exceptions” approach, since the content-based recommenders which adopted the above approach lead to dramatically decline on all metrics, except serendipity.

5.2. Managerial Implications

This research provides evidence that the presence of the “Anomalies and exceptions” approach has a positive effect on serendipity. According to previous studies, serendipitous encounter can help the user find an interesting item he might not have otherwise discovered before. Traditionally, the success of recommender systems is evaluated by predicting accuracy of recommendations. Higher accuracy of recommendations means higher user satisfaction and willingness to pay. But too obvious recommendations are not really useful because of the lack of discovery. In contrast, serendipity emphasizes the important of providing new, unexpected and desirable products to the user.

Despite the importance of serendipity, E-commerce providers should take into consideration to trigger serendipity in their recommender systems in order to satisfy the user demand of new information and unexpected products. Moreover, E-commerce providers should identify which paradigm of recommender systems is much suitable for introducing serendipity. For example, for recommender systems with less precision, collaborative-filtering recommenders are less affected by the trade-off relationship between serendipity and user satisfaction or willingness to pay, whereas content-based recommenders are affected much more than the collaborative-filtering ones.

In addition, because of the existence of a trade-off between serendipity and other metrics used to evaluate the quality of recommendations, E-commerce providers should take into consideration how to eliminate or moderate the decline of user satisfaction and willingness to pay, etc. For instance, setting a threshold to filter products among recommendation candidate pool is one possible solution. For recommender systems with high rating less precision, not only serendipity level of recommended products increases compared to classic ones, but the decline of others metrics also slight. Indeed, serendipity can be triggered through less precision approach, but the overall recommendation quality will improve based on the premise of high rating. Furthermore, E-commerce providers could attempt to adjust the less precision level and the threshold of high rating to find balance-point between serendipity and other metrics in order to provide better recommendations.

5.3. Limitations and Future Research

There are several limitations in this research. First of all, the scale and method utilized to measure serendipity was a novel method first proposed in this study. Therefore, the measure has not been proven in reliability and validity and might not truly capture the definition of serendipity stated in this study. Future researches could focus on investigating the nature of serendipity and attempt to establish the reliability and validity of the measure.

Secondly, because of the difficult to get commercial datasets, this study adopted the Netflix dataset as a basis of both the recommendation and the website catalog. In addition, DVD movie was chosen as the experiment product and online DVD rental retailer as the experiment context. Results of this study might be limited by the dataset and the experiment context that we adopted. Future research could replicate the study by using other datasets (i.e. the MovieLens dataset) and different contexts or products to generalize the results to other realms.

Finally, consumer characteristics might play crucial roles in affecting the results of user perspective on recommender systems. For example, the “effective” of recommenders might be different depending on factors such as user requirements. In fact, we considered the product awareness dimension at the beginning of the study. The definition of product awareness is the set of products that the user is initial aware of before attending to the help of any recommender system (Wu, Joung and Chiang, 2009). Consumers that belong to hit awareness are aware of relatively more hit products compared to niche products, and vice versa for ones of niche awareness. We do not discuss product awareness factor in this study because of the results between hit awareness and niche awareness made no clear distinct. This might be limited by the scale or method utilized in our study. Future researches could still focus on investigating the impact of consumer characteristics on serendipity suck like product awareness.

6. Reference

Adomavicius, G. and Tuzhilin, A. 2005. Towards the next generation of recommender systems: a survey of the state-ofthe-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 17(6), pp. 734-74

Anderson, C. (2006). "The long tail: How endless choice is creating unlimited demand." Hyperion, New York

Balabanovic M. and Y. Shoham (1997). "Fab: Content-Based, Collaborative Recommendation". Communications of the ACM. Vol. 40, No. 3.

Bawden, D. (1986). "Information systems and the stimulation of creativity." Journal of Information Science, 12, 203-216.

Bilgic, M. and R. J.Mooney. (2005) "Explaining recommendations: Satisfaction vs. promotion." Beyond Personalization Workshop, IUI.

Brynjolfsson, E., Y. Hu, et al. (2006). "From niches to riches: Anatomy of the long tail." Mit Sloan Management Review 47(4): 67-+.

Burke, R. (2002) "Hybrid Recommender Systems: Survey and Experiments." User Modeling and User-Adapted Interaction, v.12 n.4, p.331-370

Cheung, C. M. K. and M. K. O. Lee (2005). Consumer satisfaction with internet shopping: a research framework and propositions for future research. Proceedings of the 7th international conference on Electronic commerce. Xi'an, China, ACM: 327-334.

Clemons, E. K., G. D. Gao, et al. (2006). "When online reviews meet hyperdifferentiation: A study of the craft beer industry." Journal of Management Information Systems 23(2): 149-171.

DeLone, W. H. (1992). "Information systems success: the quest for the dependent variable." Information Systems Research 3(1): 60.

DeLone, W. H. and E. R. McLean (2003). "The DeLone and McLean model of information systems success: a ten-year update." Journal of Management

Information Systems 19(4): 9-30.

Fleder, D. and K. Hosanagar (2009). "Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity." Manage. Sci. 55(5): 697-712.

Ge, M., Delgado-Battenfeld, C., Jannach, D.: Beyond accuracy: evaluating recommender systems by coverage and serendipity. In: Proceedings of the 4th ACM Conference on Recommender Systems (RecSys 2010), pp. 257–260. ACM, New York (2010)

Goel S., A. Broder, E. Gabrilovich, and B. Pang. Anatomy of the long tail: Ordinary people with extraordinary tastes. In WSDM'10, NY, USA, 2010.

Good, N., J. B. Schafer, et al. (1999). Combining Collaborative Filtering with Personal Agents for Better Recommendations. Proceedings of the Sixteenth National Conference of Artificial Intelligence (AAAI-99).

Granovetter, M. (1973). "The Strength of Weak Ties", American Journal of Sociology, Vol. 78, Issue 6, May 1973, pp. 1360-1380.

Gup, T. (1997). Technology and the end of serendipity. The Chronicle of Higher Education, 44, (November 21), pp. A52

Herlocker, J. L., J. A. Konstan, et al. (1999). An algorithmic framework for performing collaborative filtering. Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval. Berkeley, California, United States, ACM: 230-237.

Herlocker, J. L., J. A. Konstan, et al. (2004). "Evaluating collaborative filtering recommender systems." ACM Trans. Inf. Syst. 22(1): 5-53.

Hijikata, Y., Shimizu, T. and Nishida, S.: Discovery-oriented collaborative filtering for improving user satisfaction, Proc. ACM IUI'09, pp. 67-76 (2009)

Iaquinta L., Gemmis M., Lops P., Semeraro G. (2008). Introducing serendipity in a content-based recommender system. Eighth International Conference on Hybrid Intelligent Systems, Barcelona, Spain. pp 168-17

Kawamae, N., Sakano, H. and Yamada, T.: Personalized recommendation based on the personal innovator degree. In ACM Recsys, pp. 329–332 (2009)

McKinney, V., K. Yoon, et al. (2002). "The measurement of web-customer satisfaction: An expectation and disconfirmation approach." Information Systems Research 13(3): 296-315.

McNee, S. M., I. Albert, et al. (2002). On the recommending of citations for research papers. Proceedings of the 2002 ACM conference on Computer supported cooperative work. New Orleans, Louisiana, USA, ACM: 116-125.

McNee, S. M.; Riedl, J. & Konstan, J. A. (2006). Being accurate is not enough: how accuracy metrics have hurt recommender systems, Proceedings of the 2006 Conference on Human Factors in Computing Systems (CHI 2006), pp. 1097-1101, Montréal, ACM

Pazzani, M.J and D. Billsus, (2007).Content-based Recommendation Systems. In "The Adaptive Web: Methods and Strategies of Web Personalization"

Oliver, R.L. (1997). "Satisfaction: A behavioral perspective on the consumer." McGraw-Hill, New York

Resnick, P. and H. R. Varian (1997). "Recommender systems." Communications of the Acm 40(3): 56-58.

Sarwar BM, Karypis G, Konstan JA and Riedl J (2000) Analysis of recommendation algorithms for e-Commerce. In: Proceedings of the 2nd ACM Conference on Electronic Commerce (EC'00). ACM Press, New York, pp. 285–295

Schafer, B., Frankowski, D., Herlocker, J., Sen, S (2007). Collaborative Filtering Recommender Systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.): The Adaptive Web: Methods and Strategies of Web Personalization. Lecture Notes in Computer Science, Vol. 4321.

Shani G. and Gunawardana A, (2009). Evaluating Recommendation Systems. Technical report, No. MSR-TR-2009-159.

Swearingen, K. and R. Sinha (2001). Beyond Algorithms: An HCI Perspective on Recommender Systems. Proceedings of the SIGIR 2001 Workshop on Recommender Systems.

Tintarev, N. & Masthoff, J. (2007). A Survey of Explanations in Recommender Systems. In G Uchyigit (ed), Workshop on Recommender Systems and Intelligent User Interfaces associated with ICDE'07.

Toms, E. G. (2000). Serendipitous Information Retrieval, Proceedings of DELOS Workshop: Information Seeking, Searching and Querying in Digital Libraries, Zurich - Switzerland

Van den Poel, D. and J. Leunis (1999). "Consumer susceptibility of the Internet as a channel of distribution." Journal of Business Research 45(3): 249-256.

Wu, L.L., Joung, Y.J. and Chiang, T.E. (2010) "Recommendation systems and Sales Concentration: The Moderating Effects of Consumers' Awareness and Acceptance." Working paper, Taipei, Taiwan: National Taiwan University, Department of Information Management

Wu, L.L., Joung, Y.J. and Lee, J.L. (2011) "The Impacts of Online Recommenders on Satisfaction and Willingness to Purchase: The Moderating Effects of Consumers' Awareness and Susceptibility to Recommenders" Working paper, Taipei, Taiwan: National Taiwan University, Department of Information Management

Yalcinkaya, G., Calantone, R.J. and Griffith, D.A. (2007). " An Examination of Exploration and Exploitation Capabilities: Implications for Product Innovation and Market Performance" Journal of International Marketing15(4):63-93

Zellweger, P. (1997). "Web-based sales: defining the cognitive buyer." Electronic Markets 7(3): 10.

Ziegler, C.-N., S. M. McNee, et al. (2005). Improving recommendation lists through topic diversification. Proceedings of the 14th international conference on World Wide Web. Chiba, Japan, ACM: 22-32.

7. Appendix

7.1. Appendix A: Measurements

A. Interest level towards recommended products
(Scale 1 to 100, 1 = Extremely Uninterested, 100 = Extremely Interested, question repeated for all 10 recommended products) 1. How interested are you with the DVD movie recommended by the recommender system?
B. Familiar with recommended products
(Nominal Scale: Yes, No and Have Seen Before, question repeated for all 10 recommended products) 1. Have you even known the DVD movie recommended by the recommender system before?
C. Overall satisfaction towards recommended products
(Scale 1 to 5, 1 = Extremely Unsatisfied, 5 = Extremely Satisfied) 1. How satisfied are you with the 10 DVD movies recommended by the recommender system?
D. Willingness to rent recommended products
(Scale 1 to 5, 1 = Extremely Low, 5 = Extremely High, question repeated for all 10 recommended products) 1. How willing are you to rent the DVD movie recommended by the recommender system?
E. Familiar with final chosen products
(Nominal Scale: Yes, No and Have Seen Before, question repeated for all final chosen recommended products) 1. Have you even known the DVD movie that you have chosen?
F. Interest level towards final chosen products
(Scale 1 to 100, 1 = Extremely Uninterested, 100 = Extremely Interested, question repeated for all final chosen products) 1. How interested are you with the DVD movie that you have chosen?

7.2. Appendix B: Participant Demographic Information

	Our Sample			TWNIC*
Variable	**Description**	**n**	**%**	**%**
Gender	Male	159	49.69%	50.26%
	Female	161	50.31%	49.74%
Occupation	Student	263	82.19%	15.74%
	Non-Student	57	17.81%	84.26%
Education	High-School	11	3.44%	30.05%
	Undergraduate	235	73.44%	24.22%
	Graduate	64	20.00%	4.40%
	Other	10	3.12%	41.33%
Income	No income	91	28.44%	31.57%
	Under 10,000	127	39.69%	9.00%
	10,001 or above	85	26.56%	55.43%
	Not sure	17	5.31%	4.00%
Age	19 years or below	53	16.56%	12.57%
	20-24 years old	196	61.25%	8.10%
	25 years or above	71	22.19%	79.33%

*: Data from the "2012 Taiwan Broadband Internet Usage Survey Report" published by Taiwan Network Information Center in March 2012