AN INTEGRATED MOBILE CONTENT RECOMMENDATION SYSTEM

Worapat Paireekreng

B.B.A. Thammasat University, Thailand

M.S. (Hons) National Institute of Development Administration, Thailand

This thesis is presented for the degree of

Doctor of Philosophy of

Murdoch University

December, 2012

DECLARATION

I declare that this thesis is my own account of my research and contains as its main content work which has not previously been submitted for a degree at any tertiary education institution.

(Worapat Paireekreng)

ABSTRACT

Many features have been added to mobile devices to assist the user's information consumption. However, there are limitations due to information overload on the devices, hardware usability and capacity. As a result, content filtering in a mobile recommendation system plays a vital role in the solution to this problem. A system that utilises content filtering can recommend content which matches a user's needs based on user preferences with a higher accuracy rate.

However, mobile content recommendation systems have problems and limitations related to cold start and sparsity. The problems can be viewed as first time connection and first content rating for non-interactive recommendation systems where information is insufficient to predict mobile content which will match with a user's needs. In addition, how to find relevant items for the content recommendation system which are related to a user's profile is also a concern.

An integrated model that combines the user group identification and mobile content filtering for mobile content recommendation was proposed in this study in order to address the current limitations of the mobile content recommendation system. The model enhances the system by finding the relevant content items that match with a user's needs based on the user's profile. A prototype of the client-side user profile modelling is also developed to demonstrate the concept.

The integrated model applies clustering techniques to determine groups of users. The content filtering implemented classification techniques to predict the top content items. After that, an adaptive association rules technique was performed to find relevant content items. These approaches can help to build the integrated model.

Experimental results have demonstrated that the proposed integrated model performs better than the comparable techniques such as association rules and collaborative filtering. These techniques have been used in several recommendation systems. The integrated model performed better in terms of finding relevant content items which obtained higher accuracy rate of content prediction and predicted successful recommended relevant content measured by recommendation metrics. The model also performed better in terms of rules generation and content recommendation generation. Verification of the proposed model was based on real world practical data. A prototype mobile content recommendation system with client-side user profile has been developed to handle the revisiting user issue. In addition, context information, such as time-of-day and time-of-week, could also be used to enhance the system by recommending the related content to users during different time periods.

Finally, it was shown that the proposed method implemented fewer rules to generate recommendation for mobile content users and it took less processing time. This seems to overcome the problems of first time connection and first content rating for non-interactive recommendation systems.

LIST OF PUBLICATIONS

[A] W. Paireekreng and K. W. Wong, "The Empirical Study of the Factors Relating to Mobile Content Personalization," *International Journal of Computer Science and System Analysis*, vol. 2, pp. 173-178, 2008.

[B] W. Paireekreng, K. Rapeepisarn, and K.W. Wong, “Personalised Mobile Game Recommendation System,” in *6th International Conference in Computer Game Design and Technology (GDTW 2008)*. Liverpool, United Kingdom, 2008.

[C] W. Paireekreng and K. W. Wong, “Adaptive Mobile Content Personalisation Using Time-of-day”, in *7th International Conference on e-Business (INCEB2008)*. Bangkok, Thailand, 2008.

[D] W. Paireekreng, “Time-based Mobile Content Usage Personalisation,” in *Postgraduate Electrical Engineering and Computing Symposium (PEECS 2008)*. Perth, Australia, 2008.

[E] W. Paireekreng, K. Rapeepisarn, and K.W. Wong, "Time-Based Personalised Mobile Game Downloading," *Transactions on Edutainment II, Lecture Notes in Computer Science*, vol. 5660, pp. 59-69, 2009.

[F] W. Paireekreng and K. W. Wong, "Client-side Mobile User Profile for Content Management Using Data Mining Techniques," in *8th International Symposium on Natural Language Processing*, Bangkok, Thailand, 2009.

[G] W. Paireekreng and K. W. Wong, “Intelligent Mobile User Profile Classification for Content Personalisation,” in the *10th Postgraduate Electrical Engineering and Computer Symposium (PEECS 2009)*, Perth, Australia, 2009.

[H] W. Paireekreng and K. W. Wong, "Mobile Content Personalisation Using Intelligent User Profile Approach," in *the 3rd International Conference on Knowledge Discovery and Data Mining (WKDD 2010)*, Phuket, Thailand, 2010.

[I] W. Paireekreng, K. W. Wong, and C. C. Fung, "Cluster Analysis for Personalised Mobile Entertainment Content," Cultural Computing. IFIP Advances in Information and Communication Technology, 333, pp. 36-45, 2010.

[J] W. Paireekreng, K. W. Wong, and C. C. Fung, "A Model for Mobile Content Filtering on Non-Interactive Recommendation Systems," in *2011 IEEE International Conference on Systems, Man, and Cybernetics (IEEE SMC 2011)*, Anchorage, Alaska, 2011.

[K] W. Paireekreng, K. W. Wong, and C. C. Fung, "Integrated Mobile Content Recommendation: A Comparison Study," in *International Conference on Internet Studies (NETs2012)*, Bangkok, Thailand, 2012.

[L] W. Paireekreng, K. W. Wong, and C. C. Fung, "A Framework for Integrated Mobile Content Recommendation System," *International Journal of Electronic Commerce Studies*. (In Press)

Summary of Publications with Respect to the Chapters and Aims of the Thesis

Chapter	Aims	Publication No
1. *Introduction*	Introduce research problems and aims	[A]
2. *Literature Review*	Present the relevant issues and arguments	[A]
3. *Establishing an Integrated Mobile Content Recommendation Model*	- Investigate and develop an integrated approach addressing the current limitation of predicting associated items including new and non-rated items for mobile users in order to enhance the mobile content recommendation system - Improve recommendations of mobile content to first time users using a client-side user's profile	[F], [G], [H], [I], [J], [K], [L]
4. *Integrated Mobile Content Recommendation: A Comparison Study*	- Investigate and develop an integrated approach addressing the current limitation of predicting associated items including new and non-rated items for mobile users in order to enhance the mobile content recommendation system - Improve recommendations of mobile content to first time users using a client-side user's profile	[F], [J], [K], [L]
5. *Mobile Content Recommendation System Using Client-side User Profile for a Revisiting User*	- Investigate and demonstrate an approach addressing the current limitation of recommendations of mobile content to a revisiting user using a self-management user profile	[F], [G], [H], [L]
6. *Using Time Context Information for Mobile Content Recommendation*	- Improve the recommendation system using dynamic time context information	[B], [C], [D], [E]

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS AND ACRONYMS

AC	Accuracy Rate
AR	Association Rules
ANN	Artificial Neural Networks
BIC	Bayesian Inference Criterion
BS	Bayesian Networks
CAR	Classification Association Rules
CBF	Content-based Filtering
CCF	Clickstream-based Collaborative Filtering
CF	Collaborative Filtering
CZCC	Cases in Zoning-Centroid for each Cluster
GI	Geographic Information
GPS	Global Positioning System
IR	Information Retrieval
k-NN	k-Nearest Neighbor
LBS	Location-based Services
MAE	Mean Absolute Error
MTCAR	Multi-level Targeting Classification Association Rule
NBTree	Naïve Bayesian Tree
POI	Point-of-interest
RI	Random Indexing
RPG	Role-Playing Game
SCM	Service Context Management

SR	Success Rate
SVM	Support Vector Machine
TAN	Tree Augmented Naïve Bayesian
TF-IDF	Term Frequency-Inverse Document Frequency
TN	True Negative
TP	True Positive

ACKNOWLEDGEMENT

I wish to express my deepest gratitude appreciation to my academic supervisor, Associate Professor Dr. Kevin Wong, for his excellent contribution to this thesis and for providing advice, assistance, and encouragement throughout this study, and especially for critically reading this thesis.

I would like to thank the co-supervisor, Associate Professor Dr. Lance Fung, for his close association and advice. My extended thanks go to the School of Information Technology and Murdoch University Library for providing facilities for my research.

I am very grateful to the Dhurakij Pundit University for the financial support and the opportunity for my study. Without this support, my PhD study would not have been possible.

Finally, I am grateful to my family for their companionship, support and inspiration along the way of the study. I apologise if I have inadvertently forgotten to acknowledge those who have assisted me in any way.

TO MY FAMILY

CHAPTER 1

INTRODUCTION

1.1 Background

The advancement of computing and the Internet have played an important role in the development of the information age. The Internet has now formed the backbone of modern communication, particularly the World Wide Web (WWW). Today, the use of the World Wide Web has moved beyond desktop computers to millions of mobile devices. While mobile devices may come in all forms, in this thesis they refer to smartphones or Internet accessible mobile phones which have incorporated many features to facilitate better information access and utilisation [1].

However, a mobile device has limitations related to the hardware, such as input capability, network constraints and memory limitation [2]. Moreover, system usability such as small screens have definite impacts on the usage and content presentation of mobile devices [3, 4].

A mobile phone can be viewed as a personal device that can provide information in many aspects. With the ubiquity of the mobile devices, a mobile Internet content provider may determine the context information of a user, and have the capability to analyse a user's profile. Such knowledge could be used to send some desirable information to a user. Nevertheless, there is the problem of information overload. With the vast amount of information available, it is a challenge to provide the appropriate information that a user really wants. In devices with higher computational power, larger memory and storage, complex computational techniques including data mining and machine learning methods could be used to match the user's need in order to provide personalised recommendations to the user [5, 6].

To address the problem of information overload, personalisation and content filtering have gained attention in the past decade [7-9]. In this thesis, the focus is limited to content filtering based on a user's needs or other influencing factors such as time and a user's demographic factors to obtain personalised recommendations. One related application of content filtering is a recommendation system which is concerned with the content being recommended, and how the contents are displayed. A recommendation system filters items from all the available items and displays only the appropriate content based on personal preference kept in the personal profile [10].

Due to the limited device capability, quality and speed of the network connection, recommendation and content filtering for the mobile content user is needed to reduce the amount of information being delivered to the user. The focus of this research is to develop techniques that can help to deliver to users, content items which match their needs from content providers. The techniques should also provide the means for lower ranking content to be presented to the user according to its relevancy against the profile of the user.

1.2 Mobile Content Recommendation

Due to the challenge of delivering a huge amount of content in a mobile Internet service, it is necessary that the mobile information page presented to be used should be appropriate according to the user's need, and other criteria such as time. This will increase the accuracy of information retrieval, and mobile content recommendation is needed to deal with this objective. This will help users to access the information they want in a shorter period of time. Users may also navigate directly to the relevant information, and the amount of information delivered to the user could be decreased.

Mobile content recommendation systems use mobile content filtering techniques to address the information overload problem by reducing the content that is presented to a mobile device user. In addition to the systems, personalised content including recommending associated or relevant items based on the criteria or conditions derived from the user's profile can be delivered to user as well. This can help a user to find the desired content based on the needs and context information such as time [11]. The mobile recommendation system addresses the problem of finding information from a large collection of information [12].

In some product marketing and sale sites such as amazon.com or movielens.org, the products or items to be recommended can be books [13], movies [14] or TV programs [5]. Most of them are based on other user's ratings or other criteria such as results from symbolic learning, user preferences or the items' categories. Such recommendation systems have also been implemented for mobile devices. This can be noted from VISCORS [15], MovieLens Unplugged [16], local-based information systems [17] or the news [18].

The work of Melville et al. [19] and Schein et al. [20] also addressed the problem of the recommendation system in terms of sparsity and cold start issues, which are due to insufficient information in making any recommendation. They, however, focused on the method and metric used for model fitting in creating the recommendation model. Nevertheless, their studies have not addressed the content and user relationship with respect to other factors, such as user demography. In addition, the technique used was based on other users' ratings instead of using the user's behaviour. The MONERS system [18] is a hybrid method used to recommend news on mobile devices using batch work mechanism and system learning which are carried out at the server. A user must also subscribe to the news service, and the user information including interests need to be submitted and stored at the server.

Content filtering in a mobile content recommendation system can be based on user profile information. Such profile information could be due to a user's input according to his or her interests, and the user's preferences, which are derived from the behavioural usage and inferred from the user's profile. The concept of a user profile in a mobile phone has been introduced by Wagner et al. in 2002 [21] and there are several mobile content frameworks such as MobileIQ [7], Lee's work [22] or the framework by Ning et al. [23] that used a user profile. The user profile has also been widely used in several applications such as Kobsa [24], Weißenberg et al. [25], Gemmell et al. [26], Castells et al. [27], Seitz et al. [28] and Jeon et al. [29].

However, using only the user profile might be insufficient to predict the user's navigation patterns and the mobile Internet usage. The situation and surrounding environment of the user are also important in determining the type of information required at the time. This can be changed or adapted according to the needs of the users. This is important because mobile information usage is based on the context information, and the important source of information required to determine the desired content is also constantly changing. This suggests that the context information, such as time, is an important factor which could be used to determine how information is delivered for the mobile devices [30, 31].

In recent years, data mining using machine learning techniques has provided more relevant information to users in mobile recommendation systems [18, 32, 33]. However, many components and stages are involved in the recommendation process in order to increase the accuracy. There is a need to search for an efficient framework that can provide content to satisfy the user's need by providing personalised content recommendation.

1.3 Research Problem

Mobile content recommendation used to provide personalised content for mobile users is a challenge, especially when a huge amount of information is available. It becomes more challenging when the context (such as time of the day) of the mobile user is taken into consideration to provide more accurate and relevant content to the user. Besides, the user's ratings may also influence what information will be displayed to the user. There is also a need to incorporate any new information or items that have not been rated previously.

However, to address the problems of such content recommendation, it is noted that such frameworks may vary due to different purposes [34]. In addition, only some mobile personalisation frameworks proposed have included mobile personalised content recommendation. Some examples are BPEL4WS [35] or mPERSONA for the portal wireless user [36]. However, to our knowledge, there is no single framework that can address all the problems together.

Content filtering is an important step in a mobile recommendation system. The main concept is to predict the items derived from other users' preferences and use the established prediction model to make recommendations to a user who has a profile close to a certain group of users. The mobile recommendation system should perform two main tasks, which are predicting the most desired items and recommending associated or relevant items based on the criteria or conditions derived from the user's profile. However, there are some drawbacks in the existing systems found in the literature [7, 34]. Most mobile recommendation systems face problems in the early stage (i.e. during the first two interactions) due to a lack of user's information. This problem is known as sparsity and cold start [19, 20] due to insufficient information being available to make the recommendation. It can be viewed in content-based perspectives as follows:

- **First time connection:** the first time a user connects the recommendation system cannot derive desired items due to the lack of knowledge relating to the user's preference. Although data mining techniques can help to solve this problem in the latter stage, to mitigate the sparsity of the first time user a model for mobile content filtering is needed.

- **First content rating:** In the early stage of new content, the amount of downloading or rating may not be sufficient to determine the item rating. Thus, some of these new contents may not be delivered to the users.

After handling the early stage problem, the mobile recommendation system should consider addressing the problem of finding relevant items for a revisiting user. In the second stage, recommendation is normally based on the preferences in a group that a user has the best fit. The group is normally established by some clustering techniques. In the next stage, after the user has provided more responses and feedbacks to the recommendation system, the recommendation will be based on the individual profile in order to achieve a more personalised content recommendation.

Based on the problems and motivations presented, this thesis focused on the development of a mobile content recommendation system for the early stage of connection or first time connection for a non-interactive system. Non-interactive here means a user does not need to input their rating, and interact with the system. In this thesis, the developed non-interactive approach is also linked to the interactive mobile recommendation system in order to incorporate users' feedback and the mobile Internet usage patterns for revisiting users.

The mobile content filtering uses time context information in order to present the most appropriate information to the user according to time. In addition, this methodology also proposes that user profiles should not remain stagnant, but adapt over time. For example, users may change their occupations and income. This reinforces the importance of using user profiles in content filtering to achieve better retrieval results.

Therefore, an integrated model that blends the user group identification and mobile content filtering for mobile content recommendation was proposed in this study, in order to address the current limitations of the mobile content recommendation system. The proposed model enhances the system by finding the relevant content items that match with a user's needs based on the user's profile. The significance of the proposed method is to improve performance of mobile content recommendation and obtain higher accuracy rate of content prediction and predicted recommended relevant content measured by recommendation metrics.

1.4 Research Aims

This research aims to:

1. investigate and develop an integrated approach of user group identification and mobile content filtering addressing the current limitation of predicting associated items, including new and un-rated items, for mobile users in order to enhance the mobile content recommendation system;

2. improve recommendations performance of mobile content to first time users using the client-side user's profile; and
3. improve the recommendation system performance using dynamic time context information.

1.5 Scope and Assumption

This thesis focuses on the mobile content filtered in a mobile recommendation system. The content preferences are based on a user's rating on the information provided, and the time context information. A prototype system is developed to demonstrate the entire framework.

The framework will operate on the assumption that users will share at least one variable related to the demographic factors. The variety of content categories will be limited to the datasets utilised in the experiments found within this thesis. In addition, the content data is rated according to the user preferences for content in terms of how much they like or dislike content, and such information is converted to a scale measurement. The user can choose not to rate the content. In this thesis, it is assumed that the communication between the mobile client and server uses the same standard. Building standard protocol for network communication is not within the scope of this thesis.

1.6 Proposed Framework

A proposed framework in this research is developed to address the problems of using the mobile Internet on a mobile device, which focuses on areas which currently are not handled by existing methodology under one framework. To overcome several problems of the device, this framework is developed with a focus on content filtering when a user connects to the Internet via the device in each session. The proposed framework can be divided into two tiers: a mobile device client layer and a content provider layer. The details are shown in Figure 1.1. The mobile client layer consists of the user profile management system for the users who are using the mobile Internet. Its purpose is to build the user profile by collecting information from the users, whereas the content provider layer refers to both the contents and services that are provided for users in order to meet their needs.

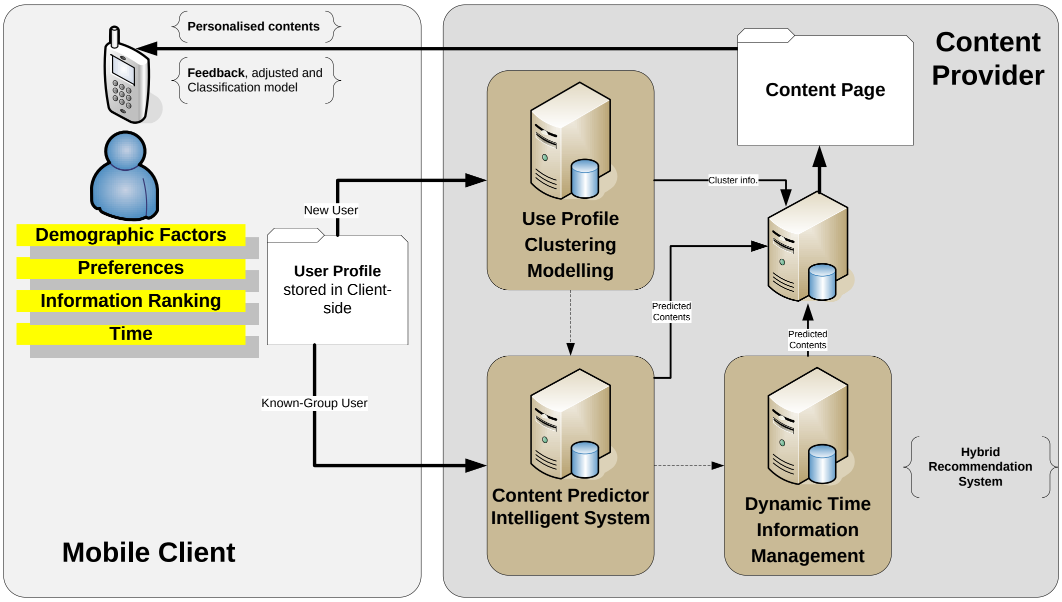

Figure 1.1. Framework for mobile content filtering for a recommendation system.

Mobile Device Client

The user profile component keeps basic personal information, such as demographic data, and tracks the user's preferences and personal information through interaction. This user profile is stored in the mobile device. Moreover, the user preferences are derived from the mobile Internet usage behaviour and feedback from the user on the retrieved information from the service provider. During the revisiting stage, the user profile will be used to make recommendations. Additionally, the user's needs and preferences may change according to the situation and context. Normally, during different times of the day, the mobile user may require different information. As a result, time is an important factor that decides what kind of information the user may want at different times.

Mobile Content Provider

As can be observed from the framework in Figure 1.1, the output that a user gets from the content provider is a personalised page generated from the recommendation system for the session that the user connects to the mobile Internet. After the previous processes, the user profile will work with the service provider's user profile system. In the user profile clustering modelling, it receives the information from the user in order to process the information of the user profile and creates clusters for the user's session when there is no user's cluster information available on the profile. It can help to mitigate the problem of sparsity for the first time connection user and for first or low rating items, and provide relevant items for the personalised content. This component applies a machine learning technique in order to construct the user's group and it also improves content recommendation for a first time user.

Processing of the component should be done offline due to high computation cost; furthermore, the mobile Internet connection and response time should be as fast as possible. Therefore, it is suggested that high computation tasks should be avoided. Consequently, the components will be separated from the online processing component. However, the content predictor intelligent system plays a vital role in predicting the possible desired information for the connected sessions, including recommending relevant items for the recommendation system. It also addresses the problem of predicting associated items for a mobile content user.

In addition, adaptability can be easily incorporated when machine learning and intelligent techniques are implemented in the personalisation process. The result from the system can also be sent to the dynamic time information management for further processing. The content will be generalised from the user profile and context information to determine the content page and data after which output from the intelligent system is combined as a content filtering to obtain a predicted set of information about what the user needs at that moment based on the context. It is organised and presented on the first page of the user's mobile Internet. In addition to adjusting the user's profile simultaneously, the feedback data that is adjusted from the intelligent system is also sent back to the mobile client in order that the user profile will be adapted consistently between the mobile and the content provider. It will update the data stored with the mobile client. Finally, the updated user profile will be used next time when the user connects to the mobile Internet, which is called revisiting. This helps to improve recommendations of mobile content for the user, as does the client-side user's profile.

1.7 Outline of Thesis Chapters

The thesis is comprised of six chapters. Chapter 2 will provide the background and literature review of mobile content recommendations and the techniques used. Here, artificial intelligence techniques suitable for mobile content personalisation will be explored as well as the integrating context information issues and challenges that a hybrid recommendation system with these techniques by means of increasing personalisation's performance. Chapter 3 will present the techniques to establish the recommendation content using data mining techniques based on a user's demographic information. This technique will be followed by discussion and comparison with other references from the literature review.

Chapter 4 presents and discusses the verification of established mobile content for the personalisation system including recommended content items as a hybrid system. Chapter 5 presents the practical mobile content recommendation system for a self-management user profile, and flow management for new users and revisiting users. Chapter 6 shows the context information of time for mobile content personalisation which is used in this study. This chapter also includes a discussion on different context information towards mobile content.

Finally, Chapter 7 presents the discussion and conclusion of the study as well as the recommendations for further work.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter reviews the current issues and research relating to mobile recommendation systems. The fundamental components of mobile content as well as the approaches used in mobile content filtering and mobile recommendation systems are examined. Data mining techniques that play a vital role in the performance of the mobile content recommendation system will also be discussed. The outline of Chapter 2 can be seen from Figure 2.1.

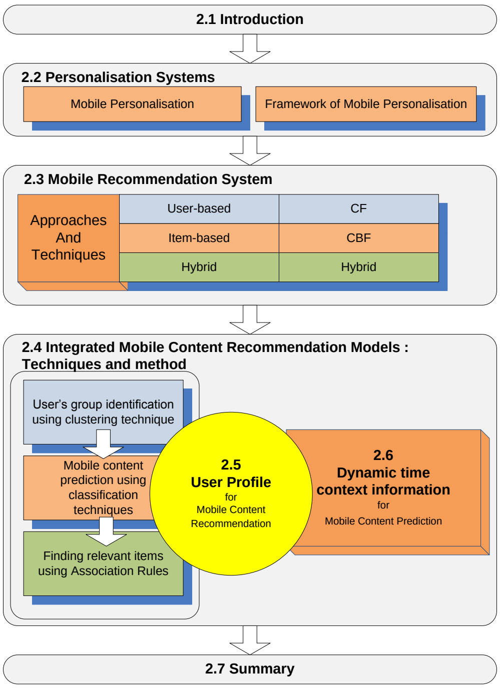

Figure 2.1. Outline of Chapter 2.

2.2 Personalisation Systems

One of the main reasons for using a recommendation system is to provide personalised content. The concept of personalisation was presented by Ivar Jorstad et al. [35] as the individualisation of something to fit a person's specific need. Montgomery and Smith [34] mentioned that most of the definitions of personalisation consists of three main elements: adaptation, individual customers, and using marketing information related to the customer. For mobile personalisation, the study by Hakkila and Chatfield [37] showed that users want to personalise their mobile phone content and their experience can be improved if the users are provided with alternative forethought [38]. In addition, despite people having a poor understanding of the extent to which phones or digital content can be personalised, the personalisation of such devices holds great importance for the users, due to emotional value [39]. These studies have shown ideas on how a mobile device is treated as a personal belonging and why people need to be served by the device individually. Consequently, the success of personalisation depends on important factors such as classification of the user profile and their context customisation [40, 41]. Content personalisation, which is itself a solution to the problem of unstructured information, can be observed in Zhang's study of mobile content adaptation [9]. Information specific to the user's needs can be provided through content filtering [42].

2.2.1 Mobile Personalisation

Surfing the Internet on mobile devices is different from surfing on a personal computer. Mobile devices have unique characteristics in terms of ubiquity, where location and time might be important in defining the user's needs in terms on how the information is used. Therefore, according to research by Zhang et al. [43], they classified personalisation for handheld devices into two main areas: content visualisation and content presentation.

Content visualisation concerns the approach to enhancing how data is viewed and analysed. For example, Borodin et al. [44] proposed CMo, a technique which can adapt original webpage content for presentation on mobile devices. Burigat et al. [45] experimented with the use of graphic maps on mobile devices and Yin et al. [46] proposed the extraction of the content of web pages in order to deliver the content suitable for the target device. It can be seen that in mobile content visualisation, a personalisation system filters information based on user participation and relevant information in order to provide personalised service.

The second area of mobile personalisation is content presentation. The research in this area generally focuses on the gathering of relevant information for presentation to the user. This approach uses user profiles and other information such as the time of the week [47] to predict the information that a user will use on mobile Internet. Some research has focused on the navigation paths that the user goes through, while others concentrated on mobile content portals which attempt to rearrange the menu on mobile phones in order to display the most wanted options in the top line of the device's screen. In addition, some research focused on the generalization of data into a universal format for all platforms. This can be observed from the majority of research related to this area, such as those in [2, 22, 48, 49]. A recommendation system is one of the applications of personalisation in this area.

To exploit the advantage of a mobile device, adaptive content, which can be adjusted when the usage changes according to the environment, becomes an important issue. Several mobile applications related to information seeking have been developed including examples such as tourist guides, news updates or classified information and services [9, 49]. The study of Feijoo et al. [50] presented the mobile content space using dimensions of mobility and creativeness. This can help to distinguish the group of contents that could adapt to mobile devices, and content that allows creativity. For example, content such as multimedia or entertainment can be classified as requiring more creativeness. Such contents need not relate to a specific mobility service, but they just adapt to a specific mobile platform. On the other hand, mobile search can be classified under both dimensions as it is a specific mobility service but also has a more creative content. This can be seen in Figure 2.2.

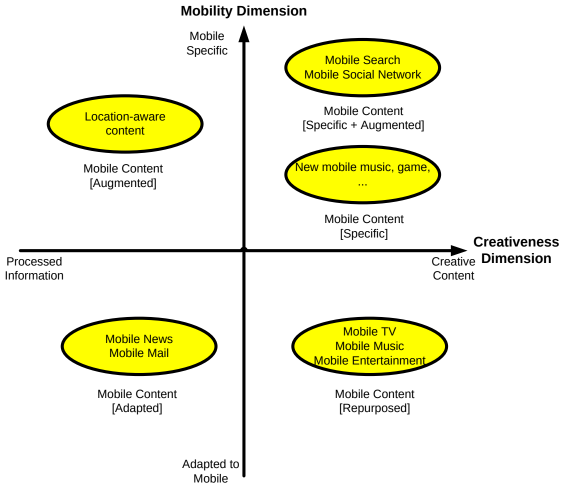

Figure 2.2. Mobile content space and categorization dimension. Adapted from Feijoo et al. [50].

Consequently, the research presented in this thesis focuses on mobile personalisation for content presentation. This is achieved through the implementation of the mobile recommendation mechanism.

2.2.2 Framework of Mobile Personalisation

The first mobile content personalisation framework was created in 2002 and has been used as a guideline for creating mobile recommendation systems. Following this, other frameworks like the MobileIQ [7], BPEL4WS [35], mPERSONA [36], Lee [22], Ning et al. [23] and Service Context Management (SCM) focused on content personalisation, while FLAME2008 [25] and Mobile GI [49] focused on location-based information system. These can be used as guidelines for establishing the recommendation system on mobile content with personalised service. The common components of the various mobile content personalisation frameworks can be seen in the summary table below, Table 2.1. However, mobile content for Geographic Information (GI) is not included because those research [25, 49] focused on location and geographical identification.

Table 2.1

Summary of Features of the Mobile Personalisation Frameworks

Features\Framework	**MobileIQ**	**BPEL4WS**	**mPERSONA**	**Lee**	**Ning**	**SCM**
Architecture	Proxy-based, Multi-agent	Web Services	Multi-agent	Multi-agent	Multi-agent	Multi-agent
Purpose	Access information	Select Web Services	Minimise click distance	Interoperation	Mobile portal personalisation	Inferring context
Application	General	General	Restaurant/ portal	Shopping Multimedia	Mobile portal	TV Content
User						
- User Model	N/A	N/A	N/A	Yes	N/A	N/A
- User Profile	Yes	Yes	Yes	Yes	Yes	N/A
- User Profile Storing	Server	Server	Client (Subscribe)	Server	N/A	N/A
- Updatable User Profile	N/A	N/A	N/A	N/A	N/A	N/A
- User Interest	Rating (1-10)	Extrinsic Intrinsic	By user	Like/Dislike by user	Yes	N/A
Context Model						
- Time	No	N/A	No	No	No	Yes
- Location	Yes	N/A	Yes	No	Yes	Yes
Intelligent System	Clustering	No	Tree	N/A	N/A	Rule-based
Privacy	Yes	Yes	Yes	N/A	N/A	N/A
Recommendation	N/A	Yes	Yes	Yes	Yes	N/A
Relevant Items	N/A	N/A	N/A	N/A	Yes	N/A

The frameworks above display certain weaknesses. The system needs the user to perform further steps, such as filling out registration or membership forms. This may unnecessarily complicate or discourage user participation as the time spent may be disproportionate to the information required. Importantly, many users fear that their personal information may be leaked as it will be stored at the content provider's server. As such, user consent should be requested prior to their information being used for the personalisation in the recommendation system. Users should be allowed to manage their own personal information including their preferred content on their mobile device.

To perform better personalised service, techniques are needed to predict user behaviour patterns. However, some frameworks either have not included the techniques or have not provided a complete solution to perform tasks such as identifying users with similar behaviours, grouping users according to their demographic factors, and making predictions about their preferences. These techniques can help content providers to improve their understanding of the users and to provide users with their desired items.

2.3 Mobile Recommendation System

The problem of trying to find information on a large database can be a difficult and time-consuming process, especially an online system. Most recommendation systems apply knowledge discovery techniques to provide personalised recommendations for information, products or services. The process of personalising content occurs when the user accesses the system. Recommendation systems not only predict the items that the users want, but will also recommend items that may be associated with the predicted item. Therefore, the main tasks of the recommendation system are to:

1) predict the target item from a pool of items, and
2) recommend the set of items that is relevant to the predicted items based on the user's preferences.

The recommendation system uses a mechanism to predict associated items and provide them to the user based on criteria such as user preferences, item popularity or demographic factors. A classical recommendation system is Ringo, which is a music recommendation system that uses other users' interest profiles to calculate similarities, and recommends items to a user who presented with similar features [51]. A prominent example of a recommendation system is the one used by amazon.com, the largest online retailer [13]. It is similar to the system used by the online auction website eBay as both utilise recommendation mechanisms to recommend products to users. Netflix and MovieLens [14] by GroupLens Research at the University of Minnesota are examples of recommendation systems that focus on finding movies based on user preferences. In addition, the recommendation system known as the Infotainment TV show has also been used for television program recommendations. The research applies symbolic machine learning to discover user interests and preferences and construct user models, resulting in the creation of decision lists [5].

Olmo and Gaudioso [10] have defined various recommendation systems in terms of user interaction. The recommendation systems can be separated into interactive and non-interactive systems. They defined an interactive recommender as a system that collects data from the user based on past interactions with the system. For example, WebWatcher requires a user to rate items in its applications. On the other hand, a non-interactive recommender gathers data before any user interaction occurs. It seeks to choose interesting or useful items for recommendation to the user. Olmo and Gaudioso note that non-interactive systems are not limited to systems that do not require user interaction. Rather, they refer to the fact that the data is collected outside the system or comes from external sources. Therefore, if the user shares some useful information, such as demographic factors, the recommendation system can utilise them to obtain the recommended items.

Thus, it can be seen that the non-interactive recommendation system should be used for the initial stages of the recommendation system where the user has not shared any information with the system. Consequently, the interactive recommendation system can be used as the next stage when the user has shared some information with the system. The recommender can then select the most suitable items for the user based on the user's information regarding content rating. However, users rarely rate the items as they generally do not wish to engage in detailed interactions and they may be concerned with privacy as well. As a result, the interactive recommendation system experiences slow growth, especially in new content areas. It requires longer time to provide appropriate recommendations. New content may not be recommended or included in recommended lists to the user despite the content being relevant to the user's need; therefore, the use of collaborative filtering and other methods is needed to solve these problems.

The recommendation system plays a vital role in mobile website browsing in order to overcome mobile device limitations due to user interface, screen, connectivity and information overload.

Many researchers have proposed personalised applications or content for a mobile user using recommendation systems. VISCORS [15], which recommends wallpapers for mobile devices to users through content filtering, is an example of such a system. Another example is MovieLens Unplugged [16], a movie recommendation system for mobile devices. This also focuses on user rating and collaborative filtering to select recommended movies. Some recommendation systems focus on location-based services (LBS) [17, 52]. These systems adjust their recommendations by using the user's short-term and long-term preferences. News recommendations [18] are another example of applications that utilise ratings-based recommendations.

Although the recommendation system is implemented on mobile devices, the main tasks of prediction and recommendation should be maintained. Those studies as shown above [15-18, 52] have focused on predictions but did not seek to create relevant items based on a user's preferences or demographic factors. Furthermore, tourism and pedestrian applications use the user's location to find useful information. The study of Zipf and Jost [49] was based on a user model related to dynamic personalised service and used adaptive GI (Geographic Information). However, the study focused on pedestrian navigation and POI (Point-of-interest) and was not concerned with user-related information or direction recommendations. It did not include product or result ratings for the recommenders and other users' opinions to construct or establish recommendation systems for the early stages.

Another problem relating to establishing the model is un-rated items or new items. New items appear on a website that have not been rated, or may have low ratings relative to older items. They will have less chance of appearing at the top of the list despite their relevance to the user. An associated problem that is faced by the recommendation system is providing content for first time and revisiting users. As a result, it is a challenge to provide personalised content for a first time user via a recommendation system.

There are two main approaches for a recommendation technique, namely collaborative filtering (CF) and content-based filtering (CBF). The basic idea is to recommend items to the target user by predicting their utility through previous ratings by other users. Typical group of users have been used as a mechanism for creating user models to compensate for the limited amount of information on each individual user.

2.3.1 Collaborative Filtering

A technique used in recommendation systems is Collaborative Filtering (CF). The main concept of CF is to use ratings given by other people. These ratings may be arranged or utilised based on specific purposes and which will then be used by the recommendation system to select items for the target user. The form of input for collaborative filtering is called a user matrix. It can be seen as user, item and rating as arranged in the matrix m x n as seen below in Figure 2.3.

User\Item	**Item 1**	**Item 2**	**Item 3**	**…**	**Item n**
User 1	R_{11}	R_{12}	R_{13}	…	R_{1n}
User 2	R_{21}	R_{22}	R_{23}	…	R_{2n}
User 3	R_{31}	R_{32}	R_{33}	…	R_{3n}
…	…	…	…	…	…
User m	R_{m1}	R_{m2}	R_{m3}	…	R_{mn}

Figure 2.3. Example of user matrix.

The process of collaborative filtering can be found in Chen and McCleod [53] and it can be summarized as:

1. collect user data, item rating and its representation,
2. find similarities of interesting subjects for the decision, and
3. compute recommendation items.

In the recommendation system, CF still requires both the user's preference and user's similarities among the user's group. Nonetheless, the recommendation system generally has problems relating to a first rater and first time connection user in terms of establishing the model for them. The challenge is to incorporate user profiles in terms of demographic factors to identify the group of users that the new user belongs to, followed by establishing models for un-rated items or new items. However, before going through those problems, it is necessary to consider the approaches in CF. Approaches used to build the CF can be divided into two main categories, namely, memory-based and model-based.

The memory-based CF uses user ratings to recommend the item based on statistical methods. It calculates similarities between a user and the item rating in the user matrix. This approach is easy to understand and effective. The example can be seen from work by Yu et al., using CF for recommender systems [54]. However, there are some drawbacks such as the scalability problem as it consumes much memory when it deals with large datasets. In addition, the size of matrix could be too big for calculation. The memory-based approach cannot handle more new users and new items due to limitation of the memory on the device and the server. As the matrix is fixed once created, it has to predict the new user's rating including new items for further prediction. This approach also depends on the user rating. If the rating is not available or insufficient, the recommendation items might not be successful.

The other approach of CF is model-based. Data mining and intelligent techniques are commonly used to establish the prediction model for recommendation. The models are created based on the training datasets. Some techniques used are Artificial Neural Networks (ANN), cluster analysis and Bayesian Networks [6, 55-57]. The advantage of the model-based approach is that it is normally more accurate, and it can handle large datasets with more flexibility. It can also deal with new user prediction [55]. The main disadvantage for this approach is the computation required to build the prediction model. An example of a model-based recommendation system can be seen from the study of Liu et al. [33]. They implemented sequential rules with CF for the recommendation system. Their approach helped to enhance the quality of the recommendation system by considering the sequence of items [58].

The other example of a model-based mobile recommendation system is related to a multiple channel product [55]. It predicts a user's desired items by combining multiple sources of information from different channels, for example, from product categories that the user bought from television, catalogues, and web channels. Furthermore, personalisation can be performed by proposing the recommended product items from the ordinary menu. Applied symbolic machine learning is used to discover a user's interests and preferences to construct the user profile implemented for an Infotainment TV show [5]. Some data mining techniques will be discussed in more detail in Section 2.3 for analysing each component of a mobile content recommendation system. This is to demonstrate how these techniques help a mobile content recommendation system perform recommendation tasks for the model-based approach.

Having considered the main approach of CF, it is necessary to look at the elements in the user matrix. The user-based and item-based algorithms are the approaches used in the implementation of the user matrix of the CF.

User-based CF assumes that many users may have the same interests. These interests can be referred to as recommending similar items to other users who may have similar interests. The technique that is widely used for a user-based algorithm is the nearest neighbour approach. It forms groups of users with similar interests in terms of ratings and makes the top N recommendation. An example of user-based CF can be found in [59].

However, the problem of sparsity, insufficient information in making any recommendation, in a user-based algorithm occurs when a user did not rate the item.

The recommendation system cannot calculate the similarity of users. For example, user A and user B who purchased the same items did not rate the purchased items. The effect of this can lead to missing information when constructing the neighbourhood and similarity index. As a result, the system will not be able to recommend the item.

User-based algorithms needs to add some factors or features such as implicit rating, implicit preference or content boost to improve the recommendation. These are reported in [60]. The explicit rating refers to rating of the items provided by users, and is included. The study also separates items into categories. It shows that this algorithm can combine the explicit rating and category-based techniques, which can perform better than conventional user-based algorithms.

Research from [44, 45] have shown that item-based CF can perform better than user-based algorithms. This could suggest that user-based algorithms by using only user similarity is not enough to provide a good recommendation.

An item-based CF approach changes the perspective of considering users' similarity to focusing on similarity between items. These items are rated by users and they will be computed by a similarity measure. Similar items in the group will be used for prediction. The example of an item-based algorithm for a recommendation system can be seen from various studies [56, 61-63].

In summary, advantages and disadvantages of user-based and item-based algorithms for CF are shown in Table 2.2.

Table 2.2

Summary of User-based and Item-based CF

Approach	Aim	Fundamental Technique	Advantage	Disadvantage
Item-based	Find k similar items that correlate rating by different user with similarity measure	Pearson Correlation Coefficient (Mapping between item pair)	Avoid computational bottleneck of user to user	Not considering user aspects such as preference and similarity
User-based	Compute correlation or similarity among users	k-nn (computing row of user matrix mapping with target user)	Take into account of the user similarity	Problem of scalability, the computational complexity increasing linearly, and the difficulty with large numbers of users such as e-commerce

Nonetheless, each approach has advantages and disadvantages. Overall, the recommendation system needs both user-based information and item-based rating information to recommend the items even for the first time user on a non-interactive system. Therefore, hybrid approaches have been designed to combine both item-based and user-based CF. For example, Liang et al. [64] added one more dimension for user and item in CF. It is the tag information which defines a user's interests and preferences. This also addressed the problem of the lack of a relationship between user and item in a recommendation system. However, this does not address the sparsity problem. Liang et al. [64] specify the number of minimum tags information stored in the experiment to avoid sparsity problem.

Other research [65] also presented a hybrid CF to handle the user and item information concurrently. This research is based on a similarity calculation using the modified Pearson Correlation by adapting user similarity and item similarity. They then added a control factor for fusing user-based and item-based algorithms. However, this method is based on Pearson Correlation, the recommendation system with a memory-based approach, where it has the problem of scalability and limitation in calculation as it relies on user or item rating. This method did not use any algorithm in finding relevant items to recommend items, where this thesis will be working to create such a method.

In most current recommendation systems, it is normally difficult to establish a prediction model of the user when they are a first time rater and first time connection user. To focus on this problem, it is necessary to incorporate an appropriate user profile, such as demographic factors, to identify the group of users that the new user belongs to. The issue related to a user profile for a mobile content recommendation system will be discussed in Section 2.4. The recommendation system needs both user-based information and item-based rating information to recommend the items, even for the first time user on a non-interactive system. However, many researchers only used item rating in making the recommendation [19, 56, 60-62, 66].

There are problems when the recommendation system depends too much on rating information. Firstly, rating information from a user dealing with many kinds of content may provide incorrect recommendations because contents may have dissimilar characteristics or a user may have different preferences. Secondly, a rating mostly uses an ordinal scale and it is explicitly collected. However, there may be a rating or content recommended based on the user's demographic information collected implicitly. Therefore, to find a group or cluster for recommendation systems, incorporating a profile based on users' information and preferences is important. Furthermore, a recommendation system should not be limited to the collection of explicit user ratings but should also include implicit knowledge for discovering user ratings or relevant content items.

2.3.2 Content-based Filtering

The other approach of finding interesting and relevant content for a recommendation system is content-based filtering (CBF). The principle of CBF comes from information retrieval (IR) and it has been used when a user needs documents or web pages. This approach focuses on analysing the content of the documents or items to be recommended to a user. In addition, it represents the most wanted item by measuring an item's importance [67]. The user profile may need to be used to filter the user's preferred content. Basically, CBF uses the preference of the current user to predict and recommend contents including unseen items to new incoming users. These items might be similar to the past rated items or have similar keywords.

The main difference between CBF and CF is that CBF focuses on finding the relationship between the content items and the user's preferences, while CF focuses on finding similarity between users' ratings or user preferred content items that are similar. The concept of content-based filtering has been used since 1995 when the World Wide Web became popular. Yan and Garcia-Molina [68] implemented a simple CBF using text filtering and profiles called SIFT on the Internet news. However, these profiles are manually constructed and updated. A user must input specific words that will be saved in the profile and must update the words manually.

Although it is a manual system, the concept that allows users to change the profile is important.

Meteran and Someren [11] found that different topics in the same document could have different precision ratios for prediction. This is because the documents contain many different terms on several topics. Besides, a term could have more than one meaning. Therefore, it is too difficult for a learning algorithm to learn the patterns and relationships between documents. Meteran and Someren [11] suggested that an effective recommendation system should be constructed based on CF and CBF jointly. By combining CF and CBF, the system can serve users who have more experience with mobile Internet content and a recommendation system. This user might have some preferences or interests that are different over a period of time or due to some special occasion. In CBF this information can be helpful for a recommendation system in case the user needs an individual preference based on his or her historic usage data. In contrast, in the CF approach, the preference of other users needs to be referred to for the user that is presently using the recommendation system [69].

Some recent work using CBF on mobile content can be observed from [70]. The system that was reported is a blog content recommender. It can identify a user's preference by counting the number of words that appeared in the documents. By using the user preference profile, the mobile recommendation system predicts the desired blog topics for the mobile phone user. The counting of the user's keywords comes from the user's browsing history and interests using the TF-IDF algorithm (term frequency–inverse document frequency), which is an algorithm used to calculate the keyword frequencies. This research applied a hierarchical clustering technique to find similar users based on a few keywords. After the group of users is found, the content will then be delivered to the users by means of push messages. However, using keywords for clustering is similar to document classification; it may not reflect the true interest of the user. Moreover, hierarchical clustering is not appropriate for clustering large data and it performs slower than other clustering techniques like k-means [71, 72].

Commonly, CBF can recommend items to the user easier when compared to CF in the case where a user profile is provided. It is able to recommend unrated items because it considers the document content when making recommendations. Besides, it can deal with new users even when a user has never used the system before, as long as the user has some profile available in order to predict his or her interests. However, CBF has some disadvantages. Firstly, it cannot filter item-based recommendations using qualitative analysis such as the statistic method or correlation analysis. Secondly, it is also difficult to separate the articles that have the same keyword with different meanings. In addition, CBF needs text attributes that are constructed by inputting to each item manually. In some research, it was found additional filtering techniques are needed on top of the CBF to address these issues [18, 70].

From CBF, it can be seen that the main components for the recommendation system are the user profile and keywords of the contents. The user profile is required to collect and record a user's preference and content of interest. It can also be used to keep track of the keywords and their frequency in order to use them as input for the recommendation system. For mobile devices, it would be better if a user can keep the keyword of his/her interests within the mobile device because it is more convenient and beneficial for the user [73-75]. One of the drawbacks of CBF is it takes a long time to learn the users' profile or to collect users' usage data at the server side. This also increases exponentially with the increasing number of users [69]. Another advantage of keeping the profile at the clients' side is that it could also increase the confidence of the users when using the mobile recommendation system, because the users have more control of their own profile. A user can manage their own profile, such as by adding more keywords or removing unused keywords, and changing their own personal information.

Table 2.3 provides the summary on the advantages and disadvantages of CF and CBF approaches for recommendation system. The table is modified from the research by Shani et al.[76] and Shardanand and Maes [51].

Table 2.3

Summary of Advantages and Disadvantages for the CF and CBF Approaches on a Recommendation System

Approach	Main idea	Advantage	Disadvantage
Collaborative Filtering	User gets advice from friends or other people	- independent from item characteristic - able to provide complicated items such as range of item attributes - dealing with any kind of content [66]	- limitation on building a good recommendation for new users with no rated items or new items—known as the sparsity and cold start problem - unable to recommend items that are dissimilar to items that were already seen in the past
Content-based Filtering	Derived concept of information filtering for documents using analysis of text/word matching with user profile	- if there is a user profile, it is easier to provide a valid recommendation to new users even when they have never used the system before - able to provide a new item that has never been rated before	- unable to filter item content in terms of quality of item or some assessment on quality including rating and ranking information - unable to distinguish similar terms in the document, and parser of text may be necessary - text/word attributes have to be assigned manually - special/add-on filtering technique is needed to address text assessment problem - need longer time to collect and learn the user preferences

2.3.3 Hybrid Recommendation System

There are advantages and disadvantages for both the CBF and CF approaches. As a result, by combining CBF and CF, the hybrid method is designed to bring the advantages of each approach to create a better recommendation system. When the CF has handled the cold start problem, the output from this approach can then be input into the CBF method. Figure 2.4 shows an overview of the approaches that are used in recommendation systems.

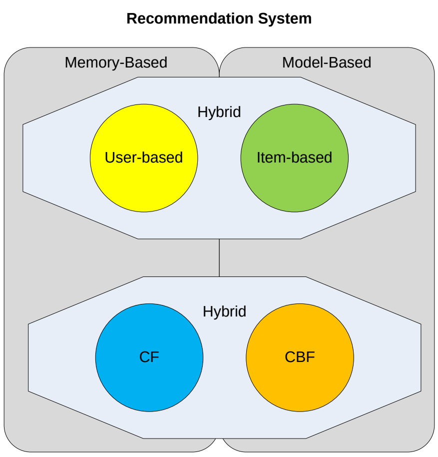

Figure 2.4. The overview of approaches used in recommendation systems.

The study by Melville et al. in 2002 [19] has shown that a hybrid recommendation system can solve the problem of rating sparsity and first rating concurrently. They created a full user-rating matrix for calculating in the CF process. The matrix is derived from the CBF process to boost rating for the sparsity problem. The CF method helps to provide recommendations for active users too. Instead of using pure CBF, in addition, CBF can be used to find other users' correlation from the user rating of the content in order for the system to generate better recommendations. Another example is VISCORS [15], the mobile wallpaper image recommendation. It used CF-based recommendations to find the similarity of the customer's neighbourhood in terms of the image rating. CBF is then implemented for the part that requires an image for recommendation. This can help the user to reduce search time for the desired image. Schein et al. [20] worked on a hybrid recommendation system. They addressed the cold start problem by using content data, the description and detail of data associated with an item, for an unrated item. However, this study focused on the method and metrics for model fitting.

The work described so far only makes use of the rating information from the user. It does not include a user's demographic factors in finding specific preferences.

A news hybrid recommendation system that combines CF and CBF known as MONERS was reported on by Lee and Park [18]. Their recommendation system showed the read ratio on the news. From the experimental results, it appeared that the read ratio from the recommended news menu is higher than from the category of the news menu. This is suitable for content that has several pages. However, if it considers the number of news services accessed by the menu, the number of categories of news read is higher than the recommended news. MONERS uses recency of time on news in establishing the model-based recommendation system. It creates its own model for delivering news services. By using the time factor, it can address the sparsity problem as well. This recommendation system also uses user demographic factors in making any recommendation. Basically, it clusters the users based on their demographic factors and makes the assumption that users in the same group have similar article usage patterns.

Although the MONERS tries to complement personalised news, learning or analysis are done as a batch work. It does not provide a user profile at the client side, but a user must subscribe to the news service and the information including interests will be stored at the server side. In addition, this model is suitable for documents that concern recency and present time of information like news, despite the fact that preferences may not need recency or categories. A user can have more than one specific interest in different topic and different categories. Table 2.4 shows the example of a recommendation system and applications and Figure 2.5 represents the positioning of these applications towards a recommendation approach.

Table 2.4

Example of Mobile Recommendation Systems and Applications

#	Area	Application	Researchers
A	m-commerce	Product	Liu and Liou [55]
B		Product	Liu et al. [33]
C		Product	Linden et al. [61]
D	Media Content	General	MediaScout [76]
E		Music	COFOSIM [66]
F		Image	VISCORS [15]
G	Content	Blog	M-CRS [70]
H		News	MONERS [18]
I		TV Program	Infotainment [5]
J		Home improvement	PRES [11]

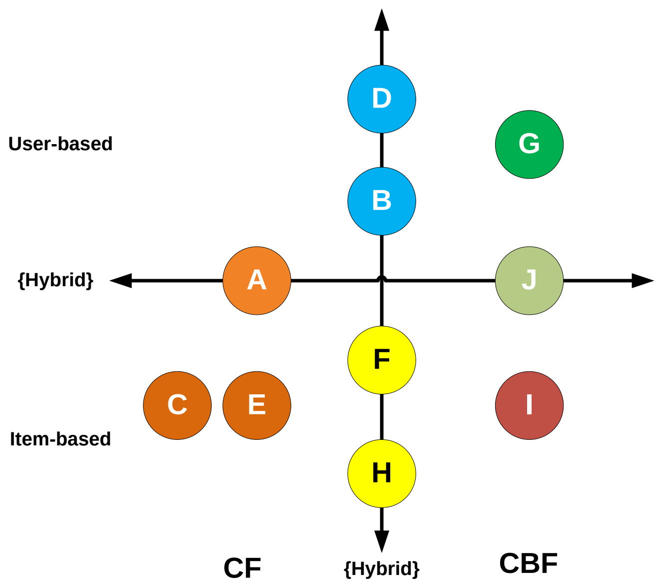

Figure 2.5. Research position for recommendation systems according to Table 2.4 information.

As a result, both CF and CBF seem to contain important features that require constructing a better recommendation system. CF is needed based on a user's preferences and similarity with the group, and provides recommendation for first-time users. Once, users have interacted with the system they might have more preferences that are different due to the period of time or on special occasions. Content-based Filtering (CBF) should be bundled in the recommendation system for the next phase to recommend content to match a user's individual preferences. It is therefore recommended that the hybrid approach should combine the features of CF and CBF together to fulfil the recommendation process for the first time connection and revisiting connection processes.

In summary, the information sharing between a user and the recommendation system is important, because it affects the items recommended which could match the user's need. Figure 2.6 shows the summary of relationship between recommendation and other factors.

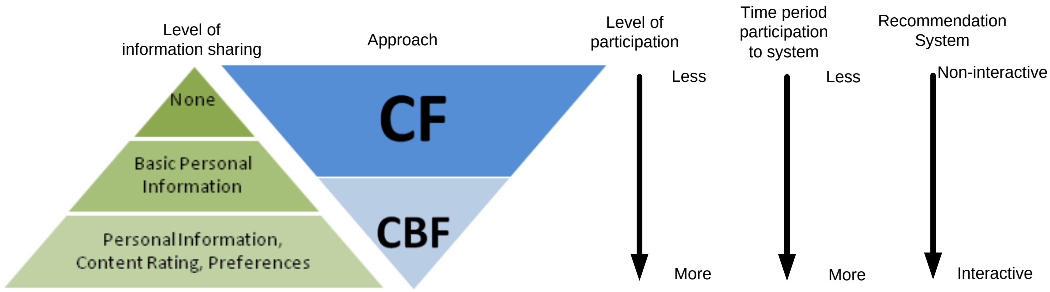

Figure 2.6. Relationship between recommendation system approaches, information sharing and related factors.

2.4 Integrated Mobile Content Recommendation Models Using Data Mining

Data mining is used by some recommendation systems to find the relationship among the users and items before making recommendations to users. There are many data mining techniques that can be used to discover the knowledge from a large database, such as clustering, classification and association rules (AR). Data mining in the area of machine learning finds the relationship among items [77-79]. Data mining using AI techniques has been applied in many applications. Intelligent agent using AI is embedded in computer especially PDAs, like e-mail and news filtering [80].

From the review, typical integrated mobile content recommendation models are observed and summarised in Figure 2.7.

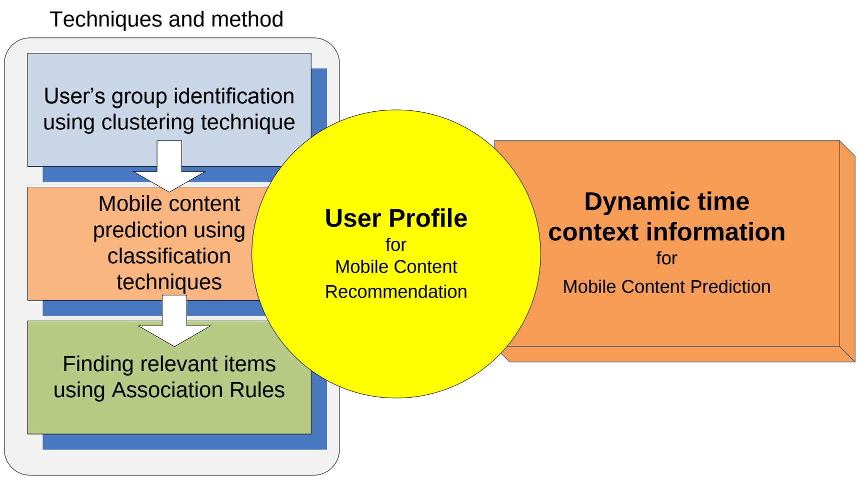

Figure 2.7. Components of typical integrated mobile content recommendation models.

2.4.1 Data Mining and Mobile Content

There are some works using CF or combination models to obtain better efficiency and accuracy of prediction. For example, Fink et al. [81] proposed the use of a user modelling server to provide content for news channels in order to create a one-to-one relationship. The system used multiple learning components, such as a user-learning component and CF, to assign a user to a peer group. Smyth and Cotter [82] used a probabilities approach for the menu manager and profiles by storing a list of access in hash tables keyed on menu ID to perform profiling and information filtering tasks. These are used to learn the users’ preferences in order to adapt the structure of the

WAP portals menu and reduce a user's navigation time. Furthermore, a hybrid model for presenting mobile content is presented, such as clickstream-based CF [59]. The hybrid mode used four methods which are the Markov model, association rules, sequential association rules, and the default model of clickstream-based collaborative filtering (CCF). The paper showed that the hybrid model performed better than other models. Likewise, Zhang and Jiao [83] used association rules to predict semi-structured data in B2C e-commerce applications. They suggested that information overload could be overcome by this technique. The model learned knowledge from historical data and combined a set of associated text documents with frequently occurring text patterns in order to recommend products for a user on a mobile phone.

The next section will discuss some common data mining techniques that were used in mobile content recommendations, specifically clustering, classification and association rules techniques.

2.4.2 Clustering Used in Mobile Content Data Mining for Mobile User Group Identification

From the industry point of view, amazon.com mentioned that a clustering model for recommendation systems is important [61]. Clustering will be broadly applied on recommendation systems to target markets in order to identify the group for a new user. In addition, they suggested that a clustering model allows better scalability online and performs better than a traditional collaborative filtering technique (CF) [61]. Furthermore, Chen and McLeod [53] mentioned the problem of CF relating to clustering, which is how a user should be dealt with and allocated to a cluster.

Machine learning plays an important role towards business data processing, especially in clustering [84]. Wu et al. [85] have shown that some commonly used algorithms in data mining include k-means, SVM, Apriori, PageRank and Naïve Bayes. They also described k-means as a simple iterative clustering method. As this is a simple algorithm with very different adaptation according to the applications [86], the clustering component described in previous studies [87, 88] also showed how the clusters are constructed based on mobile user clustering using demographic factors and information ranking. Therefore, clustering is an important stage in the mobile recommendation system because the target market needs to be known for recommendation analysis.

The research on the behaviour of the mobile Internet user can be observed from Yamakami's research [72]. Yamakami used the aging analysis model to identify mobile Internet user behaviour. This model used statistical techniques to divide the users into four groups based on the amount of access time of the users. From the evaluation results, it is shown that the prediction accuracy of user presence in the next month for mobile Internet usage is around 84-87%. The author [72] also suggested that users can be separated into heavy, common and light users based on their user experience in terms of the length of using the mobile Internet. This research focused only on the frequency of the user using the mobile Internet.

In 2006, Okazaki [89] included attitudinal and demographic information in his research by considering both continuous and categorical variables in the cluster analysis. He determined the number of clusters to be four clusters based on Bayesian

Inference Criterion (BIC) techniques. He separated the demographic profiling and attitudinal profiling before analysis. After that, the Two-Step algorithm was used for clustering based on the distance measure of the categorical attributes. Furthermore, research by Yamakami [90, 91] developed a formula to identify the appropriate number of clusters using a method known as 'revisit ratio', comparing two, three and four clusters. These researchers use the k-means clustering technique. Their results showed that cluster results have a positive relationship with the revisit ratio based on content freshness.

In addition, Yamakami [91] also presented how the system could take in the information of a user's clickstream to know how mobile Internet users spent their time in each time zone (Always On, Prime Time and Irregular Time). In another study [92], the authors aimed to extract the knowledge from the information of the customers and understand what additional services each cluster of customers would adopt. They used factors analysis relating to call usage, payment behaviour and additional service usage. After that, clustering analysis using k-means combined with association rules were implemented. The results showed that there were eight clusters which can be grouped into three main groups, specifically, new generation (leisure and entertainment usage), low-priced or fare data service, and general usage (no additional services). Although that research implements similar techniques to this thesis to identify a customer group, it focused on finding the pattern of a customer's mobile service usage behaviour, whereas in this thesis association rules are used to find additional mobile services based on the identified characteristics of each cluster from primary clustering results.

It can be observed that most of the researchers have focused on the clustering in terms of mobile Internet user behaviour, such as adoption or experience in the mobile Internet [72, 89-92]. There were only a few researchers that investigated how to determine the number of clusters and the use of clustering techniques for a mobile recommendation system [93-95]. In addition, due to the limited computational resources on mobile devices and the user need to get the response as fast as possible it is always a challenge to build an efficient system. K-means and its acceptable k-values, which are simple, are normally used for unlabelled mobile content user clusters [71, 72, 96].

2.4.3 Review of Classification Techniques for Personalised Mobile Content

Classification is related to identifying what class new observed data belongs to. Each classification technique can handle different kinds of variables and situations. An example of this is the Decision Tree. This technique has advantages as it produces an understandable map and requires low computational processing [97]. It can also handle numeric and categorical data, whereas some techniques work well only with one type of variable.

There are researchers, such as Nurmi et al. [98] or Cufoglu et al. [99], who focus on mobile personalised content classification, and their research provides a comparative study of different classification techniques for different applications. Nurmi et al. [98] presented comparative analysis of different classification algorithms. They divided the experiment into two groups: Group 1 and Group 2. Group 1 consisted of random tree and TAN (Tree Augmented Naïve Bayesian) while Group 2 used

Bayesian Classifier and JRip (Rule inducing using Ripper algorithm). In the experiment, the attribute category was separated by time, GPS, and physiology. The measurements for these comparison studies are accuracy rate and mean absolute error (MAE). There was strong assumption in the experimental data that all attributes are independent of each other. They showed that Group 1 performed better than Group 2 because its general characteristics can induce the hierarchical context data attributes and these attributes seem to be correlated with other attributes. It also works well across users. However, the predictor in this group is quite straightforward and has a more user-friendly interpretation that means ease of use to even a novice user. It can be seen that the contextual information for a mobile application in this experiment is based on absolute GPS address and there is no user information involved with personalisation. In addition, the research is focused on person by person, and a small group of users were part of the experiment.

However, there is a research related to classification techniques comparison which focused on the user profile [99]. This study aimed to find the best classification algorithm with user profiling measured by accuracy rate and the time consumed for building the model. The research used 18 attributes related to the user profile, such as age, working class, occupation, gender, interest in music, interest in books, et cetera. The results show that Naïve Bayesian Tree (NBTree) is the best classification technique with the highest accuracy rate and lowest error rate. However, this algorithm requires a longer time to build the model when compared to those methods used in the comparison. In addition, this research provides personalisation application in terms of general use without any other preferences or context information.

In addition to general classification techniques, in 2007 Kotsiantis [100] presented the comparison of supervised machine learning classification techniques. The researcher shows that the algorithm selection stage to find the classifier is important. The key points from the research also show that the importance of selection of the classification algorithm is not concerned with superiority compares to other classifiers but it should consider the conditions of a specific method towards application problems for significant performance. After that, an ensemble of classification techniques can fulfil the complement of each technique's weakness and strength.

From this research it is evident that the classification techniques can be integrated to increase the power of prediction by consideration of the features and instances of available datasets, and how the combined classification model will add or remove some of them properly.

There is a study of mobile personalisation for mobile services selection by Mahmoud et al. [6]. This study attempts to increase the accurate service selection using a feed-forward back-propagation neural network, as a classifier based on context information. It is incorporated to assist the selection of different types of particular mobile services. It uses some forms of classification on all the available mobile services. The research suggested that selecting the best available service is not a simple task. They identify the system success rate (SR), which is derived from correctly classified services and to services, at around 87%. However, this study focuses only on selecting the services without any recommendation or providing relevant services.

2.4.4 Association Rules for Mobile Content Recommendation

Association rules (AR) is one of the rule-based techniques that was proposed by Agrawal et al. [101]. It works with large transactions in a database to find the interesting relationship between data, and constructs the rules. The model starts with a set of items and there are transactions in the database which contain sets of items. The association rules can measure the quality of the rules by two metrics, support and confidence. The important feature of the AR is completeness, which is searching all possible rules based on determined support and cofidence level while the other techniques can only find the subsets of those rules. A well-known algorithm is Apriori presented by Agrawal et al. in 1993 [101]. The data format is an item matrix that has binary value in each cell that means a purchase or no purchase. Apriori is a fast algorithm although there are lots of numbers of items in databases [102]. AR can be applied in several applications. For example, Kumar and Reddy [103] merge AR to help the classification problem by rule pruning to improve classification accuracy. The other example in the real world is e-commerce using associative classification techniques on websites to recommend products such as mobile phone models to users [83].

There is also research related to personalisation and CF for recommendations that are item-based. Ye [104] proposed the concept to use association rules combined with a self-organising map. It attempts to address the problem of recommendation systems in terms of data sparsity by filling the vacant, where necessary, using the rules. However, this study is not concerned with finding relevant items but is focused on finding target items to target users, and the results are not presented yet. The other example is Mobasher et al. [105]. They implemented association rules for a personalisation system to provide better performance of the CF based on the clickstream of users and webpages. The research focused on the computational advantage over the k-nearest neighbour technique. But, they ignored the problem of the determining and clustering of the user groups. In addition, even though users gain specific content matched to their needs, they may need other content related to those specific ones. Nevertheless, those researchers did not cover mobile content nor user related issues of finding associated items.

The research by Sohn and Kim [92] implemented AR to find the additional mobile services. They extracted knowledge from the customers to understand what additional services each cluster will adopt. The minimum support is set at 9% and confidence level is set at 80% to find the rules. This research focused on forming the group of users with additional services but not finding the relevant services to present to a user. Association rule has been used in mobile applications to find the top N items as well. For example, Liu and Liou [55] find recommendations for mobile users by multiple channels weighting. Another work focused on the segmentation of users with the k-nearest neighbor method for CF. It implemented association rules to find the top N number of items based on customers' content usage behaviour (recency, frequency and monetary) [33]. However, when association rules alone are used in the recommendation systems for mobile content recommendation, it may require a significant amount of computation to find all the possible rules. Alternative approaches are therefore required to speed up this process.

After that, adaptive association rules are applied for rule extraction. The solution for this problem can be used for partitioning and targeting. Partitioning can help to reduce the number of itemsets to be counted, rather than dealing with all the items in the entire database [79]. The Classification Association Rules (CAR) is an alternative method for this approach. However, this technique can only be used for problems with the known-class database. So, the classification phase is necessary for this method. The multi-level association rules [79] technique is another adaptive association rules technique which divides the problem into levels for extracting the rules. It is a hierarchical concept in which the higher levels of frequent itemsets have more support than the lower levels. The minimum support in the same level is identical. The advantage of this method is no complete rule processing is required as the frequent itemsets in the higher levels helps filter the itemsets in the lower levels with less minimum support. This means the lower levels needs less support to run the algorithm for rule extraction and it will be run within the frequent itemsets of the higher levels. This could significantly save computational time in extracting the association rules.

To consider the usefulness of association rules towards content personalisation, there is a lot of content on the content provider's server which is updated and added to frequently. In addition, these contents have not been downloaded due to newness and their rating not being high. However, a user may need them for up-to-date news, while they have little time to search the new content because of the limitation of the mobile Internet and their available time. As a result, this research will use the association rules to address those problems by proposing relevant content to the user.

2.5 User Profile on the Improved Mobile Content Recommendation Systems

This section reviews how user profiles can be used to establish a model for mobile recommendation systems. It also presents how a user profile can improve mobile content recommendation.

The concept of a user profile was introduced by Wagner et al. in 2002 [21]. The research proposed a framework for advanced personalisation of mobile services using a profiling technique that utilises the semantic enrichment service. The main concept assumes that the user may belong to a specific group. Thus, a generalised usage pattern can be applied for that group. In addition, important standardisation efforts can facilitate online security for identity purposes. The profiling technique is device dependence, which is compatible to mobile terminal such as CC/PP and UAProf [106]. An open mobile network using XML interfaces [106] is a representative tool for connecting with other systems to facilitate this as well.

Kobsa in 2001 [24] presented the development of the generic user modelling system. It described the characteristics of a generic user model which mainly consisted of generality including domain independence. In addition, the user profile should be universal. Although the generic profile is created for one application, it can normally be used generally for other applications as well. The research also suggested that the generic user model can be applied for mobile devices in the future. The construction of the user model was later expanded to mobile user behaviour modelling [107] which was based on a task-oriented model using ontology. This approach implemented a generality concept in task-oriented problems, such as buying a book and entering a park, and scalability towards several providers.

As can be seen, if the user modelling can use the combination of relevant tasks and domains, it can provide better personalisation. User modelling can help to construct a user profile with demographic factors for determining a user's characteristics. This can provide a group of a user and recommend an appropriate content match to a user's needs. A user profile also helps content providers to serve various types of mobile content users' needs according to their profiles and to enhance the mobile content recommendation system. The user profile can also address the problem of early stage in the recommendation system which is related to the lack of a user's information. In this case, more customised content and services can then be launched. However, those researchers [21, 24, 107] mentioned above did not cover what kind of information in a user profile could be used to identify a user's group and a user's preferences. It also did not indicate the limitations of the recommendation system in terms of a first time user, first content rating or handling a revisiting user. However, there is research on the mobile user profile, which is the mobile profile based on distributed grouping (MoPiDiG) [28]. It is a location-based information system which was developed to define local grouping and distributed grouping by using user profiles. However, it focused on forming a group of users which have similar profiles in order to send a customised message. In addition, this research did not cover the finding of relevant content items for a first time user and a revisiting user. It also cannot serve the first time user or content without rating information properly.

As explained above, other parties can use this useful information to process appropriate recommendation content to a user. A greater sharing of information will lead to better levels of personalised recommendation from other parties. Therefore, to enhance the mobile content recommendation system, a user profile should be used as it allows sharing of information to produce more accurate recommendations to the users.

2.6 Dynamic Time Context Information for Mobile Content Filtering

This section describes how dynamic time can be used to improve the mobile content recommendation system. It can be observed that using only the user profile might be insufficient to predict the user's navigation patterns and mobile content usage. Having considered that mobile devices are used in a predictable manner and the user's travel patterns are normally consistent, the conditions of the information usage can be changed to adapt to the needs of the users. This is important because information usage is based on the surrounding environment and the context, and the important source of information required to determine these is also constantly changing. Incorporating context information in a mobile content recommendation system enhances the system's performance in terms of higher precision and accuracy of content prediction.

2.6.1 User Navigation for Mobile Content Filtering

The World Wide Web on the mobile Internet is normally displayed as a menu list or options. A user has to click or choose the desired option to go to the next menu or page. Those sets of options or menus are called user navigation. The research on predictive web navigation should be linked directly to the mobile devices, as different devices have different hardware and usability characteristics. For example, Buchanan et al. [108] proposed three different styles of display for news headlines, relating to the mobile's usability. As can be observed from the content presentation or content summarisation on the mobile Internet, the content personalisation might be focused on predicting user navigation. The aim of this research area is trying to reduce the click distance from the mobile Internet usage session between the first option menu displayed and the desired option menu. This also addresses the problem of content filtering. It tries to display relevant information topics which are needed by a user quickly [3, 47, 82, 109, 110].

Therefore, the beginning of mobile content prediction has been done on user navigation prediction. However, the structure of navigation might be complicated and it cannot predict the dynamic webpage in which data and information are retrieved from the database. The URL of the link is dynamic and sometimes it cannot be kept as a link in the web browser history. Hence, user navigation prediction should be focused on the most content that user want regarding to user's ambience information. Moreover, beyond this point, the recommendation system has played a vital role in predicting the desired content for a user, and it also suggests the relevant items that a user might need at that time. The example of context information about time that is added in user navigation prediction can be seen from Halvey et al. [111]. In addition, some work of the navigation prediction concentrated on the computation to obtain a possible user click menu. Some favourable models which are applied to solve the problem are the Markov model, the Bayesian network and Naïve Bayes [47, 59, 112-115]. Probabilities with hit table which included the user navigation information from the server log are also presented in ClixSmart as a user navigation method [82]. However, these researchers mentioned above did not cover the early stage problem of the mobile recommendation system and they did not integrate a user profile approach for collecting user's information to enhance the recommendation system.

In some of the work mentioned above, it was suggested that an individual model was not sufficient to predict user navigation. Therefore, a hybrid model or a combination with other information could enhance the performance [47, 112, 114, 116]. Research by Dong-Ho Kim et al. [59] also suggested that hybrid models are better. Some research focused on rearranging the page results or menu in order to provide a user with an alternative menu on the mobile device screen [109, 117]. However, the research on prediction user navigation did not take the prediction of relevant information and the user's information in a user profile into consideration.

Characteristics such as time context information for predicting the information to be delivered according to a user's situation are also not used. Moreover, most of them use Naïve Bayes as the classification method. This is mainly because the knowledge of the classes is available despite the unknown class' data needing other techniques to complement this issue. This is, thus, one of the main focuses to be considered.

Generally, context information can be separated into low-level and high-level context information. It depends on how difficult it is to determine this information. For example, a user's absolute location using GPS can be defined as high-level context information. A user's current time which may be defined by a server can be described as low-level context information. However, if the system needs more specific time matching with a user's activity, it would be more complicated. Therefore, it might be between low- and high-level context information for this case. Although the research of McMullan and Richardson [118] mentioned mobile phone usability in various aspects of its usage by assembling added hardware such as a touch screen or GPS, most people are concerned about their privacy. They do not want to be located by the GPS [119]. Therefore, the low-level context information such as time could be an appropriate factor to be integrated into the recommendation system. This should enhance the mobile recommendation system.

2.6.2 Time Information for Mobile Content Recommendation

One of the important pieces of information which is related to content presentation is time. This can help a user to reduce the clicks on their mobile device or filter content based on time. It can help a user reach the desired content quickly, which in term can enhance the recommendation system. Time context information can be classified by weekday or weekend with the information of the time-of-day. For example, context was described as a description of a situation similar to case-based reasoning when used in a mobile [48, 120]. The research suggested that system adaption is required for the change in the environment. Contextual information played an important role towards information retrieval in successive searches. From Wang and Shao [121], it can be seen that using time-framed information can improve the performance in the prediction of the future browsing patterns. This research used time-framed separation of week and semester information by using association rules. They also suggested that a user would prefer different kinds of information during different times of the day. In this thesis, models that deal with these suggestions will be addressed.

However, these researchers did not cover finding relevant items according to time, the first content rating and un-rated content. User information and user profile were not used as well. In addition, determining groups of users based on a user's preferences and content usage can be also investigated. Therefore, the integrated approach to enhance the mobile recommendation system is needed.

2.7 Summary

Mobile content filtering for a recommendation system can be classified as a non-interactive and an interactive system. A non-interactive system focuses on a new user coming to the system at an early stage. The user needs to have recommended content items matched to his or her user profile and demographic factors. A model-based recommendation system and collaborative filtering approach were discussed.

They are the common methods that have been used in the mobile recommendation system. The disadvantages of each approach, mainly cold start and sparsity, were described. This leads to the motivation of this thesis to address the limitations by establishing an integrated mobile content recommendation system. Existing mobile content recommendation did not cover determining the mobile content user group or finding relevant content items. Common techniques, specifically clustering, classification and association rules in data mining, for establishing the integrated model were discussed. It also aims to solve the current limitation of the existing system, which is the problem of the first time user and first content rating.

After that, an interactive recommendation system focuses on a user's sharing information, the user profile and a revisiting user who has been repeatedly using the recommendation system. This system encounters the problems of slow growth in new content areas and requires longer time to provide an appropriate recommendation match to users' needs. This also includes finding relevant content items. Hence, the method to handle this revisiting user is a concern. Content-based filtering and the user profile for mobile content recommendation systems were discussed. The integrated approach to handle a revisiting user and content-based filtering were investigated.

In addition, the personalisation concept has been incorporated in order to obtain personalised content to users. Mobile content filtering using time information is also implemented to find the appropriate content regarding time. This is to enhance a mobile recommendation system.

CHAPTER 3

ESTABLISHING AN INTEGRATED MOBILE CONTENT RECOMMENDATION MODEL

3.1 Introduction

This chapter proposes an integrated mobile content recommendation model using a user's group identification, and associative techniques, to address the current limitations of mobile content recommendation systems. The proposed model will address the problem of a first time user and first content rating in the early stages of the system. The system will also predict relevant items for mobile content users to enhance the system's performance in terms of content prediction. In this model, techniques will be proposed to enhance the extraction of knowledge from a mobile user profile and content profile for building the mobile content recommendation system. In order to solve the early stage problem of the mobile content recommendation system, clustering techniques will first be used to group the mobile content user for a user-based system. Following that, mobile content prediction will use classification techniques to predict the target items for an item-based concept. After that, association rules are implemented to find relevant items, which could be new or un-rated.

3.1.1 The Mobile Content Recommendation Model

The work flow of the integrated mobile content recommendation model is shown in Figure 3.1. The first function begins with user group identification. It processes the user's information in order to identify the mobile content group the user belongs to.

Next, mobile content filtering will be performed. The top ranking content will be predicted based on the user's group information. After that, the association rules that relate to the predicted items from the previous stage and the user's group will be generated. Then, the recommendation generator will collect predicted items from the mobile content filtering component including relevant items based on the generated rules. The final recommendation is generated and sent to the user.

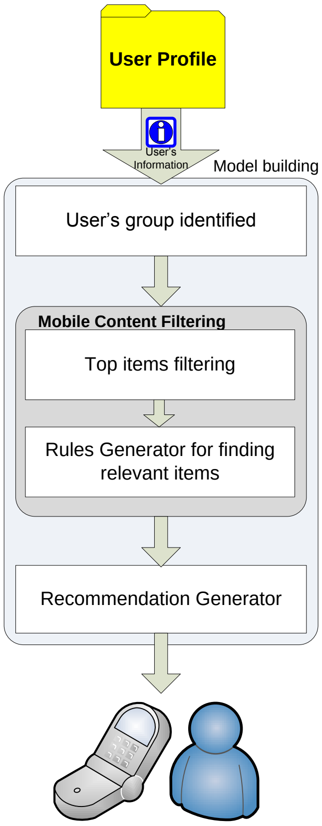

Figure 3.1. Diagram of building integrated mobile content recommendation model.

3.1.2 K-means and Determining the Number of Clusters

The k-means clustering technique is a popular technique used in mobile user group identification. This section will examine some background of the k-means clustering technique. The k-means technique has been introduced by Tou and Gonzalez in 1974 [96]. It uses a partitioning technique to cluster the data and it is a flat clustering method. K-means can handle large amounts of data and cluster faster than many other techniques, like hierarchical clustering [71, 72]. Although there are many usages with k-means in areas such as image processing or information retrieval [72, 122, 123], simply due to its implementation and speed specifying the number of optimal cluster seems to be a challenge. Much research addressing this challenge has been done.

Tibshirani et al. [93] presented the ‘gap statistic’ to determine the number of optimal clusters which can be implemented on any clustering technique, such as k-means or hierarchical clustering. This research focused on well-separated clusters and uniform distribution datasets. It calculated within-cluster sum of squares around the cluster means. In 2009, Muhr and Granitzer [94] proposed automatic cluster number selection by applying x-means with the split and merge cluster method. They measured cluster validity with BIC and F-Score. The results showed that it can select cluster numbers with acceptable runtime even for online update data on information retrieval (IR) datasets. This technique is appropriate with known class or labelled data. Another cluster number determination method is the ‘L method’ [95] which is based on the ‘knee’ or point of the maximum curvature. However, this method did not work well with global evaluation metrics and it is unable to work with applications where less numbers of clusters are found.

Cluster evaluation. Ray and Turi [123] proposed a method to evaluate the clustering technique using k-means. It is a validity ratio which is defined as:

$$validity = \frac{Intra}{Inter}.$$

Intra refers to intra-cluster distance between a point and it cluster centre. Inter is the distance between clusters. The validity is computed to evaluate a quality of clustering in terms of well-separated cluster. The concept of this measurement is minimising the sum of squared distance for intra-cluster and maximising inter-cluster value. If the validity value is small, it can be inferred that the cluster is compact compared with other k-values. This evaluation metric has also been used by Hussein [124] in the mobile recommendation system focused on system architecture. In addition, Hussein also uses k-means for clustering.

Therefore, to consider the quality of the cluster with an unlabelled cluster, the fundamental of clustering related to distance of cluster should be maintained. Firstly, the data in the cluster should have less distance from the centre of the cluster for minimising intra-cluster value. It can be implied that the data distribution in the cluster is good. Secondly, the distance between clusters should be maximised. The higher distance between clusters represents that it is good clustering. This basic concept can be used as the fundamental to measure the quality of unlabelled or unclassified cluster.

3.1.3 Labelled Data for Clustering Problem

F-Score has been used Muhr and Granitzer's research [94] to measure the quality of cluster. However, to measure the quality of clustering, one should know the class label of the cluster; otherwise, it may not show the quality of cluster properly. This problem is similar to Random Indexing (RI) which requires some testing data such as true positive (TP) or true negative (TN) to calculate the measurement. Although, there is research using fuzzy C-means clustering which proposed ‘induced entropy’ to evaluate the cluster, it also needs testing data for known classes such as visited and recommended web pages in order to compute relevance, completion and bad recommendations for presenting cluster quality [125]. The authors suggested that the focus should be on the quality of clusters produced rather than the computational complexity of the clustering algorithm.

As described above, the first stage for establishing the model of a recommendation system for a mobile content user is a user group’s identification. This is to address the current limitations of the system which are known as the first time user and first content rating. It can be observed that most of the researchers have focused on the clustering in terms of mobile Internet user behaviour, such as adoption and experience of the mobile Internet. In this research, the clustering to be incorporated in the model to address the problem in the first stage of the recommendation system is extended.

However, there is another problem that challenges the clustering for mobile content user identification for the recommendation system. It is to determine the optimal number of clusters. In clustering mobile Internet user behaviour, there was also research on the number of clusters, but this research worked well with known class data. They are not suitable for unknown class data. Therefore, the method to identify the number of clusters for unknown class data for a mobile content user group is needed. In this research, Zoning-Centroid is proposed to find the optimal cluster numbers for mobile content clusters in the recommendation system. This will be discussed further in Section 3.2.

3.1.4 Content Prediction for Mobile Content Filtering in the Recommendation System

Classification techniques have also been incorporated into the available mobile services. Mahmoud et al. [6] have suggested that selecting the best available service is not a simple task. For example, Artificial Neural Networks (ANN) with a feed-forward back-propagation neural network was incorporated to assist the selection of different types of particular mobile services [6]. Other research by Cufoglu et al. [99] had proposed which classifier is the most appropriate for classifying user profiles in the same way as Nurmi et al. [98]. Their work also presented the analysis of personalisation techniques for contextual data. However, these proposals have not taken care of the problem in the early stage of the mobile content recommendation system for first-time users and un-rated or new items. It mainly focused on the prediction and the accuracy for the identified or known classes. Furthermore, they were not used to handle the relevant items, which not only are new but could also be rated lower for content recommendation. Some classification techniques are suitable for specific kinds of user data [99] and when they are combined with other techniques, some missing information may not be used to continue with the next phase of the recommendation system in finding the recommended or relevant items [100]. Hence, a more appropriate classification method needs to be established.

Association rules (AR) have been used in mobile applications to find the top N items, as well. For example, Liu et al. [55] used association rules to find multiple channel recommendations for mobile users using channel weighting. Another work focused on the segmentation of users with the k-nearest neighbor method for collaborative filtering [33]. Using association rules alone in the recommendation system for mobile content recommendation may require a significant amount of computation to find all the possible rules. Thus, alternative approaches are required to speed up this process.

As described above, one of the aims of this research is to develop a model such that it can address the current limitations of mobile content filtering for the first time user by incorporating classification techniques. This can help to predict top content items based on user group characteristics. After that, the problem of first content rating and finding relevant items can also be addressed by AR. It can enhance the mobile content filtering performance by recommending more relevant items to mobile content users. This will be discussed further in Section 3.3.

3.2 Extracting Mobile Content User Based on User Profile Using Cluster Analysis

This section presents the mobile content user group identification with proposed Zoning-Centroid clustering.

3.2.1 The Proposed Zoning-Centroid Clustering for User Profiling on Determining Cluster Number

An evaluation method called Zoning-Centroid is proposed for the purpose of finding a clustering number for a mobile content cluster for a recommendation system. The distance from the centre of each cluster should be used to determine the cluster's members in each cluster, and to ensure that they are appropriately distributed. This method will help to select the appropriate number of clusters to establish a mobile content recommendation model with reduced calculation for the number of data.

Zoning-Centroid will use the distance between the centre of the cluster and data to calculate the zone that this data is sought in. It measures the distance from the centre of this data. The zone will be divided into five zones. Each zone is computed from *Zone-Distance* which is derived from the difference between the maximum distance in the cluster and the minimum distance in the cluster:

$$ZoneDistance_{(n,i,k)} = (MaxDistance_{(n,i,k)} - MinDistance_{(n,i,k)}) / 2^n$$

where n=zone number of cluster i and k = k-values; $1 \le n \le 5$; $1 \le i \le 5$; $4 \le k \le 8$.

Then, the *Zone-Limit* will be calculated from *Zone-Distance* as follows.

$$ZoneLimit_{(n,i,k)} = ZoneLimit_{(n-1,i,k)} + ZoneDistance_{(n,i,k)}$$

where n = zone number of cluster i and k = k-values; MinDistance = n-1 for n=1 where n=1 is referred to Zone 1 which is started from the centre of the cluster .

After that, the distance of each data will be assigned to each zone according to its limits. For example, if Cluster 1 and Zone 1's limit is 0.924737, the data with distance below this limit will be in Zone 1. In contrast, if the distance of the data is over that limit, the data will be assigned to the subsequent zone. Figure 3.2 shows the concept of zoning.

The amount of cases that fall in each zone can be measured. The count of the number and percentage in each zone is used to determine the data distribution based on the concept of Zoning-Centroid. This evaluation method will be applied for k-means between k=4 and k=8 for mobile content usage.

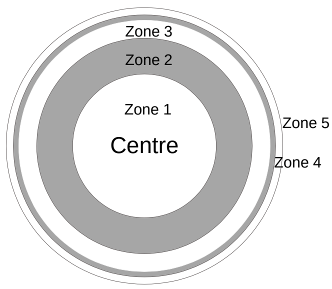

Figure 3.2. The Zoning-Centroid diagram shows the data coverage for each zone.

The Zoning-Centroid separates each zone using exponential distance from the centre of each cluster. The main concept is based on proximity between data and the centre. The first zone normally consumes 50% of the distance from the minimum to half of the maximum distance after which the second zone expands from first zone's boundary until half of the remaining data, another 50% of the overall data, are covered. This second zone consumes 25% of the overall data. Therefore, the first two zones cover 75% of the data. The objective of defining typical values is to determine the data boundary when data is irregularly distributed, specifically, as the distance between data increases when data is far from its centre. It is similar for the clustered data distribution. This is because a good quality cluster should contain data close to its centre as much as possible. Therefore, the first zone covers an area larger than the next zone. In contrast, if the zone division implements linear zoning, the distance in each zone would be separated from each other equally. As a result, the data will be difficult to separate as most data will appear to be near to its centre.

3.2.2 Experiment Results for Identifying Mobile Content User

The data source used for the experiment was obtained from the published research on the mobile Internet content users in Bangkok [126]. This set of data consists of the users' content preference, such as multimedia, and news or information services on the mobile Internet. Three hundred randomly selected records were used as training data for clustering. There are several factors and attributes in the dataset. In this research, the key demographic factors of gender, age, income and occupation have been selected to find potential groups or clusters. These attributes were chosen in acquiring the requisite data from the mobile Internet users, as well as for the ease of classification for further analysis.

The cluster analysis is performed using the k-means cluster technique. The k-means clustering technique was selected as it provides a simple algorithm that can be used to determine cluster sizes. And as has been reviewed earlier, it is a more popular clustering technique used so far in mobile applications. This allows the implementation of a clustering model at the server of the content provider in order to know the customers' characteristics and provide appropriate content to each cluster based on the cluster characteristics.

The aim of the experiment is to analyse the group based on demographic factors. The analysis should generate the appropriate number of clusters for mobile content users, leading to the identification of contents which these clusters of users will be accessing. The experiments are conducted with k-means where k=4, 5, 6, 7 and 8 consecutively.

3.2.2.1 Number of Clusters Using Zoning-Centroid

As can be seen from Table 3.1, with the number of cases in Zoning-Centroid for each cluster (CZCC), and as expected, the percentage of cases that fell in Zone 1 was the highest percentage in each k-value. The cumulative percentage of cases between Zone 4 and Zone 5 was around 5-8%. Referring to Table 3.1, these numbers were calculated by the sum of CZCC-Zone 4 and CZCC-Zone 5 values from k=4 to k=8. It can be implied that 92-95% of data approximately had not fallen over to Zone 3 for every k-value. In addition, it showed the highest percentage was in Zone 1 followed by Zone 2 and Zone 3, which meant the data for k=5 was disseminated appropriately, especially in the first two zones. The percentages and trends of each k-value and CZCC in each zone were shown in Figure 3.3 and the cumulative percentage of CZCC was shown in Figure 3.4.

To consider the cumulative percentages of dissemination of data compared to percentile of distance from the centre of the cluster to its limit, the data from Zone 1 to Zone 4 are summed together, which are 93.75% of the percentile of distance as a typical value. This shows that the actual data distribution from k=4 to k=8 is at around 94-96% in the cumulative effect of the four zones. The highest was at k=4 and it had decreased slightly when k-values increased. However, the percentage rose again when k=8. When compared with cumulative percentages of the three zones, the result still presents trends similar to four zones. These results were shown in Table 3.2.

K=5 showed the cumulative percentage from Zone 1 to Zone 2 at approximately 94% which was significantly higher than the other k-values. The percentage comparison can be seen from Table 3.2. According to CZCC, it seems that k=5 showed the most significant result compared to other k-values based on Zoning-Centroid consideration with less cumulative zones (two zones). This can be implemented to choose the appropriate number of clusters for mobile content usage.

Table 3.1

Number of Cases in Zoning-Centroid in Each Cluster

Cluster	# Cases Zoning-Centroid	k=4	k=5	k=6	k=7	k=8
1	Zone 1 limit	58	22	66	25	9
	Zone 2 limit	21	21	0	2	4
	Zone 3 limit	3	0	2	0	0
	Zone 4 limit	0	1	0	0	0
	Zone 5 limit	2	1	4	3	1
2	Zone 1 limit	44	29	20	20	20
	Zone 2 limit	9	10	4	4	0
	Zone 3 limit	2	1	2	2	1
	Zone 4 limit	1	1	1	0	0
	Zone 5 limit	5	3	1	1	1
3	Zone 1 limit	57	66	33	67	43
	Zone 2 limit	20	0	2	22	25
	Zone 3 limit	4	0	0	0	7
	Zone 4 limit	2	0	0	0	0
	Zone 5 limit	3	3	1	11	4
4	Zone 1 limit	61	47	47	12	12
	Zone 2 limit	7	16	16	3	3
	Zone 3 limit	0	0	0	3	3
	Zone 4 limit	0	0	0	0	0
	Zone 5 limit	1	6	6	1	1
5	Zone 1 limit		68	43	63	63
	Zone 2 limit		4	0	6	6
	Zone 3 limit		0	1	0	0
	Zone 4 limit		0	7	0	0
	Zone 5 limit		1	3	1	1
6	Zone 1 limit			38	11	22
	Zone 2 limit			1	3	9
	Zone 3 limit			0	0	2
	Zone 4 limit			1	0	0
	Zone 5 limit			1	1	1
7	Zone 1 limit				27	19
	Zone 2 limit				9	13
	Zone 3 limit				2	0
	Zone 4 limit				0	1
	Zone 5 limit				1	2
8	Zone 1 limit					24
	Zone 2 limit					1
	Zone 3 limit					1
	Zone 4 limit					0
	Zone 5 limit					1
	Total	300	300	300	300	300
	CZCC - Zone 1	73.33%	77.33%	82.33%	75.00%	70.67%
	CZCC - Zone 2	19.00%	17.00%	7.67%	16.33%	20.33%
	CZCC - Zone 3	3.00%	0.33%	1.67%	2.33%	4.67%
	CZCC - Zone 4	1.00%	0.67%	3.00%	0.00%	0.33%
	CZCC - Zone 5	3.67%	4.67%	5.33%	6.33%	4.00%
	Total	100.00%	100.00%	100.00%	100.00%	100.00%

Table 3.2

The Percentage Sum of Data Dissemination in Various Zones

Zone	K				
	4	5	6	7	8
Sum 4 Zones (93.75%)	96.3333%	95.3333%	94.6667%	93.6667%	96.0000%
Sum 3 Zones (87.5%)	95.3333%	94.6667%	91.6667%	93.6667%	95.6667%
Sum 2 Zones (75%)	92.3333%	94.3333%	90.0000%	91.3333%	91.0000%

* In bracket means the percentile of distance from centre to its limit and it is cumulative of typical value

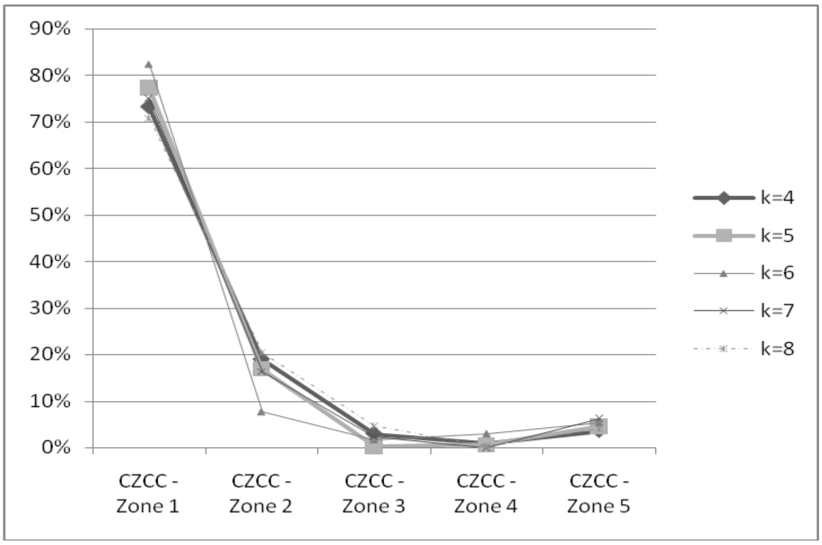

Figure 3.3. Percentage of cases of Zoning-Centroid in each zone in each cluster.

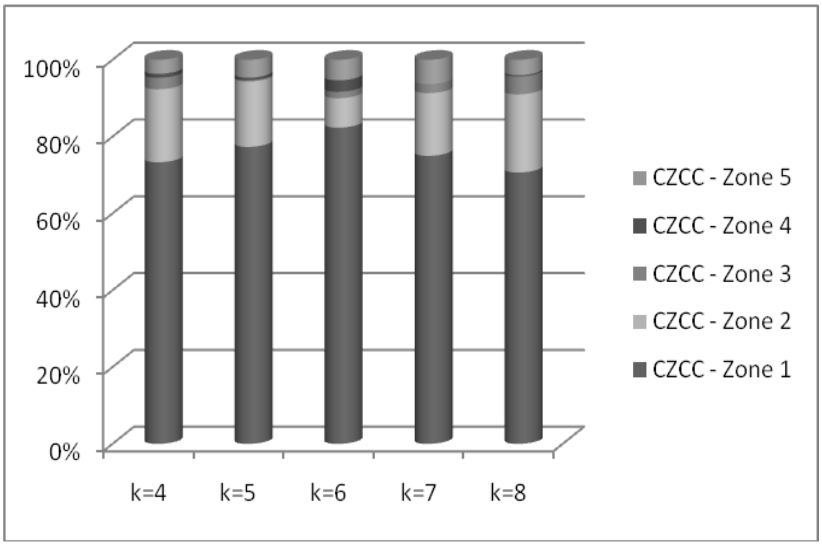

Figure 3.4. Cumulative percentage of cases of Zoning-Centroid in each zone in each cluster.

3.2.2.2 Cluster Evaluation for Mobile Content Filtering

To evaluate the quality of clustering, a method to measure the number of clusters is used and the results are shown in Table 3.3. Two-Step clustering techniques are used to compare the results of the number of clusters. The Two-Step clustering algorithm is appropriate for clustering large data and needs rapid procedure to form clusters. The algorithm is also suitable for categorical and continuous data [127]. This method was used in Okazaki's research for determining the number of clusters and mobile internet adopter cluster solutions [89]. However, the Two-Step algorithm with BIC (Bayesian Information Criterion) and ratio of distance measure showed that the number of auto-clustering for this mobile content usage dataset is just two clusters for the same dataset mentioned at the beginning in Section 3.2.2 [126]. As a result, the BIC measurement will be ignored because the results for the clustering are unable to be implemented in the further stage, such as a customer's pattern of content usage. It is too small a number of clusters for providing different content of a mobile content user group. The proposed method can show most numbers of clusters compared reasonably to auto-clustering with Two-Step.

Table 3.3

Two-Step Auto-Clustering

Number of Clusters	Schwarz's Bayesian Criterion (BIC)	BIC Change	Ratio of BIC Changes	Ratio of Distance Measures
1	3159.024			
2	2685.935	-473.089	1.000	2.309
3	2526.266	-159.669	0.338	1.099
4	2388.193	-138.073	0.292	1.560
5	2328.365	-59.828	0.126	1.004
6	2269.064	-59.302	0.125	1.291
7	2241.147	-27.917	0.059	1.117
8	2224.555	-16.592	0.035	1.094

The quality of clusters was then measured. As in this case, unlabelled data with the fundamental concept of clustering to measure the quality was used. It is the validity metric from Ray and Turi [123]. The concept is based on the measurement of the minimum distance within clusters and the maximum distance between clusters. The results were presented in Table 3.4 as follows:

Table 3.4

The Validity of Clustering

k	Intra	Inter	Validity
4	499.4445	3.8780	128.7895
5	430.2511	3.9670	108.4574
6	372.2088	3.2829	113.3793
7	360.6455	2.7056	133.2981
8	298.2944	2.3483	127.0281

According to the validity metric, the intra-cluster value was calculated from the sum squared of the distance in the cluster while the inter-cluster value was selected from the minimum value of distances between cluster centres, what was desired to be maximised. The validity could imply that if the value was small, it meant the cluster quality was good. From the results, it implied that for this mobile content clustering problem, the appropriate number of clusters should be five because its validity was the smallest compared to other k-values' clustering.

To consider the concept of Zoning-Centroid, it can be seen from Table 3.5 that this method could reduce the number of data or cases to be calculated for finding the number of clusters, and at the same time ensuring the quality of cluster by at least 5.6667%. It also decreased the percentile of distance to be considered to 75% from the cluster's centre.

Table 3.5

The Comparison of Validity and Zoning-Centroid using Cluster k=5.

	Normal Validity	Zoning-Centroid	% of calculation reduced
Number of cases	300	283	5.6667%
Percentile of distance	100%	75%	25%

3.2.2.3 Analysis of Cluster with k-values

The characteristics of each cluster based on demographic factors and content usage were analysed and the detail was shown in Table 3.6. This was concluded by using different k-values as follows;

k=4. The results showed that gender and age do not have any effect on clustering except on Cluster 4. It showed that the unique characteristics are teenager, low income and studying. For income and occupation, they were different in the three clusters. Therefore, they were unable to be determined precisely.

k=5. At this k-value, demographic factors and income started to show some significance and separated more precisely. In addition, age had a clearer influence on the cluster than the previous k-value. Teenager was still the dominating attribute in clustering while there was no effect to clustering with gender.

k=6. The cluster of teenager was maintained and gender still had no effect towards clustering. Age and occupation seem to be clearer. There were different ages in each cluster, such as above 18 years old, above 36 years old or between 19-35 years old. In addition, income had begun to be separated into less than average income and above average income.

k=7. There was one cluster where a proportion equal to 5% appeared, that is, in Cluster 6 and age began to influence the clustering. Income is then clustered more precisely in Clusters 1 and 2 with lower income and higher income groups. Occupation showed the groups which have free time and are employed with a low income in Clusters 3 and 4. Similarly with the other k-values, the teenager cluster was separated clearly compared to other clusters.

k=8. The k-value was stopped at this point by setting up a cut-off point when the small cluster has a proportion less than 5%. The teenager group together with the gender group had an effect in Clusters 1 and 8 by providing a division between the male and the female with different combinations of ages. Furthermore, occupation also determined the group characteristic by presenting as employed or having more free time. In Clusters 3, 4 and 7, where there were ages between 19-35 years old, it showed a difference by occupation and income between the clusters.

Table 3.6

Cluster Characteristic Compared to Various k-values

Cluster	1	2	3	4	5	6	7	8
k=4								
Gender	M,F	M,F	M,F	M,F				
Age	>18	>18	>18	<25				
Occupation	1,2,3	2,3,4,5,6	4,5,6	1				
Income	< 20K	>15K	<30K	<10K				
Cluster Labelling	Employed, less income	Mature, more income	Free time, more income	Teenager				
k=5								
Gender	M,F	M,F	M,F	M,F	M,F			
Age	19-45	>36	19-45	<25	19-45			
Occupation	2,3,4	3,4,5,6	2,3	1	4,5,6			
Income	>15K	>10K	<15K	<10K	<15K			
Cluster Labelling	Employed, mature, income	Free time, mature	Employed, less income	Teenager	Free time, mature, less income			
k=6								
Gender	M,F	M,F	M,F	M,F	M,F	M,F		
Age	>18	>36	19-45	<25	19-35	>36		
Occupation	1,2,3	2,3,4,5	2,3,4	1	4,5,6	3,4,5,6		
Income	<20K	>15K	>15K	<10K	<15K	<30K		
Cluster Labelling	Employed, less income	Mature, more income	Employed, mature, more income	Teenager	Free time, mature, less income	Free time, mature, more money		
k=7								
Gender	M	M,F	M,F	M,F	M,F	M	M,F	
Age	>36	>36	19-45	19-35	<25	>36	19-45	
Occupation	2,3,4,5	2,3,4	2,3,4	5,6	1	4,5,6	2,3,4	
Income	<15K	>15K	<15K	<15K	<10K	10K-30K	>15K	
Cluster Labelling	Mature, less income	Employed, mature, income	Employed, mature, less income	Free time, mature, less income	Teenager	Male, free time, more income	Employed, mature, income	
k=8								
Gender	M	M,F	M,F	M,F	M,F	M,F	M,F	F
Age	>36	>36	19-35	19-35	<25	>26	19-35	>46
Occupation	4,5,6	3,4	3,4	5,6	1	2	2,3,4	4,5
Income	10K-30K	>20K	<15K	<15K	<10K	10K-30K	>15K	<15K
Cluster Labelling	Male, free time, more income	Mature, high income	Employed, mature, less income	Free time, mature, less income	Teenager	Business owner, mature, more income	Employed, mature, more income	Female, aging, free time

* Occupations are student, business owner, employee, government officer, retired and other.
** Income is in Thai Baht.

3.2.3 A Verification of the Established Model for a Mobile Content Recommendation System on Clustering Analysis

To verify the method of Zoning-Centroid in identifying the numbers of clusters for selecting the best user clusters, experiments were conducted for another two datasets. Dataset B and Dataset C have been created by randomly selecting 400 records from the primary data to obtain each dataset. After that, each dataset was also randomised again in order to separate the dataset into training data and testing data with the same proportion as a first dataset (Dataset A), which is 300 and 100 records respectively.

When the two datasets were obtained, data distribution has been analysed as well. It appeared that data distribution of the three datasets had the same distributions in each variable which would be used for cluster analysis, specifically, gender, age, occupation and income. After distribution and variables proportion were analysed such that each dataset had a similar distribution, the process of cluster analysis has been performed until the validity of clusters were obtained.

The results as shown in Table 3.7 showed that Dataset A and Dataset C presented similar results, with the best valid number of clusters being five, while Dataset B showed the number of clusters at eight based on the smallest validity compared to other k values.

Table 3.7

The Comparison of Validity of Three Datasets

k	Dataset A Validity	Dataset B Validity	Dataset C Validity
4	128.7895	103.5026	123.0579
5	108.4574	102.1014	112.9808
6	113.3793	124.2385	121.3621
7	133.2981	95.3933	126.4155
8	127.0281	94.3327	141.5085

When the clustering has been done, there could be other problems. Outliers is one of the problems [79]. The outliers can be described as errors of data, and sometimes are considered as correct data but which present much different features from the majority of the data. Therefore, when clustering is performed, removing outliers ensures that clustering can perform better.

For Dataset B, because the number of clusters is different from the other datasets, the number of members in each cluster has been analysed. It can be seen that there are small clusters happened after the clustering process from k=6 until k=8.

Table 3.8

The Number of Members in Each Cluster in Each k-value for Dataset B

Cluster	k=4	k=5	k=6	k=7	k=8
1	74	73	74	73	27
2	54	52	66	4	35
3	107	52	81	87	22
4	65	79	32	27	31
5		44	43	35	3
6			4	48	73
7				26	27
8					82
Total	300	300	300	300	300

In this experiment, the outliers were described as a small cluster that had been clustered from the clustering algorithm and which affected the number of clusters for this method. The small cluster was determined at 5% of the total number of data. It was based on the statistic of normal distribution of the value for the number of members in each cluster. If the number of members in that cluster is less than 5% of the total record, it would be counted as a small cluster or could be classified as an outliers cluster.

Finally, the small cluster would be removed or merged with a larger cluster. According to Dataset B, although the result from the first pass of the cluster determination showed that it is eight, but in k=8 there is a small cluster appeared in

Cluster 5. In the next smaller validity step, with k=7, there was still a small cluster that appeared in Cluster 2. By repeating the process, it was found that a reasonable validity was at k=5 as there was no small clusters or outliers cluster. As a result, the best number of clusters for Dataset B should be five.

3.2.4 Discussions on Determining Mobile Content User Clusters for Establishing an Integrated Mobile Content Recommendation Model

The experiment described in the previous session not only recommended the optimum number of clusters for mobile Internet content user groups but also provides the techniques to cluster through the use of k-means and subsequent evaluation with Zoning-Centroid. The clustering generated is based on demographic factors with the data provided by the users allowing the cluster analysis to be processed easily. The Zoning-Centroid can assist in determining the appropriate k-values for the number of clusters, allowing the content providers to focus on individual clusters and deliver the right content to the right group at the right time. This can help the first stage of establishing an integrated mobile content recommendation model.

The results of this section potentially increase the value of the recommendation system by determining the optimal number of clusters to be grouped for mobile content recommendation. The appropriate number of clusters is determined by the combination of a clustering technique with fundamental demographic factors. This can form group of user to provide mobile content suitable for each cluster in the early stage of mobile content recommendation system when there is insufficient information regarding the user.

The k-means is a simple algorithm, and therefore is suitable to be used for the mobile content filtering to fulfil personalisation. The integrated model can be built at the content provider's server and can predict the user's group from the incoming user's profile faster. This addresses the limitation of a mobile content recommendation system in the early stage for a first time user and a new user. The system can proceed to the further stages of establishing the integrated model. When the content provider knows the user's characteristics, it would be easier to provide appropriate content to them quickly based on their user profile.

3.3 Mobile Content Filtering for the Recommendation System

After the mobile content user analysis is completed, the next phase is mobile content filtering for the user. This phase will proceed from the results in the previous phase to predict the desired items for the user in each cluster. The assumption for prediction in different clusters is based on the user's variety of preferences. As a result, the classifying of mobile content to serve the user's needs is important. This addresses the limitation of the early stage problem in most mobile content recommendation systems. In this phase, it can be divided into two main tasks which are items prediction and finding relevant items.

3.3.1 Mobile Content Top Items Prediction

This section concerns mobile content filtering in terms of top items prediction using classification techniques. The classification results of mobile content filtering on the top items prediction and classification model selection are shown.

3.3.1.1 Methodology and Experiment on the Top Items Prediction Using Classification Techniques

From the datasets used in previous experiments related to a mobile content user, the data has been separated to do the classification experiments into six groups which are an unclustered group and clustered groups numbered 1 to 5. After that, the top three mobile content items in each group are calculated based on the average score. The top three highest scored items have been chosen to work with in the classification experiments.

The target variable is the item that users may need for their connection session. Before feeding the data into the classification model, all data and variables have been normalised. In addition, as the binary classification technique is used, the target variable which is the user's preference rating (1 to 5) is converted to binary (0 and 1) for the prediction, where 0 is derived from the user's preference range from 1 to 3 while 1 is derived from the rating of 4 or 5. So, 0 means the user is not interested in this item, whereas on the other hand 1 represents the user's preference towards this item. This methodology is derived from the binary prediction process in which a user considers the prediction of 4 or higher in a 5-point scale rating as a desired item [56],

[20]. The experiment was carried out based on technique by technique and item by item, that is, the first item, the second item and the third item consecutively.

The classification techniques that are used in this experiment were Artificial Neural Networks (ANN), Support Vector Machine (SVM), Bayesian Networks (BS) and Decision Tree with C5.0 algorithm. These are chosen because they are popular classification techniques.

3.3.1.2 Classification Results of Mobile Content Filtering on the Top Items Prediction for Recommendation Systems

Three datasets are used in this experiment. They are Dataset A, Dataset B and Dataset C. Each dataset consists of users' information related to demographic factors, clustering information and target items' rating. These variables are used to conduct the experiment of mobile content filtering on the top items prediction. The results of the accuracy rate of top 3 mobile content prediction based on user profile and preferences towards items were shown in Table 3.9. Table 3.9(a) to Table 3.9(c) showed the results in each dataset from Dataset A to Dataset C respectively. Figure 3.5 to Figure 3.7 have shown the accuracy rate for the first, second and third item compared to each dataset and each classification technique. In addition, Figure 3.8(a) to Figure 3.8(c) have shown the accuracy rate for the first, second and third item compared to each cluster for all datasets respectively. The results of the different classification techniques can be shown as follows:

Table 3.9

The Accuracy Rate of Top 3 Mobile Content Prediction Based on User Profile and Preferences Towards Items

Table 3.9(a) Dataset A

Item #	unclustered			
	NN	SVM	BS	C5.0
1	64.0000%	63.0000%	62.0000%	64.0000%
2	66.0000%	66.0000%	61.0000%	62.0000%
3	59.0000%	62.0000%	58.0000%	64.0000%

Item #	Cluster 1			
	NN	SVM	BS	C5.0
1	75.0000%	75.0000%	83.3333%	75.0000%
2	66.6667%	66.6667%	58.3333%	66.6667%
3	66.6667%	75.0000%	75.0000%	66.6667%

Item #	Cluster 2			
	NN	SVM	BS	C5.0
1	60.0000%	50.0000%	60.0000%	50.0000%
2	70.0000%	50.0000%	70.0000%	80.0000%
3	50.0000%	30.0000%	0.0000%	50.0000%

Item #	Cluster 3			
	NN	SVM	BS	C5.0
1	69.6970%	63.6364%	42.4242%	66.6667%
2	60.6061%	63.6364%	27.2727%	72.7273%
3	57.5758%	54.5455%	54.5455%	66.6667%

Item #	Cluster 4			
	NN	SVM	BS	C5.0
1	80.7692%	80.7692%	80.7692%	84.6154%
2	61.5385%	61.5385%	61.5385%	69.2308%
3	69.2308%	69.2308%	69.2308%	69.2308%

Item #	Cluster 5			
	NN	SVM	BS	C5.0
1	57.8947%	47.3684%	63.1579%	47.3684%
2	36.8421%	47.3684%	42.1053%	42.1053%
3	36.8421%	36.8421%	42.1053%	47.3684%

Table 3.9(b) Dataset B

Item #	unclustered			
	NN	SVM	BS	C5.0
1	64.0000%	60.0000%	56.0000%	56.0000%
2	54.0000%	62.0000%	56.0000%	63.0000%
3	53.0000%	58.0000%	54.0000%	56.0000%
Item #	Cluster 1			
	NN	SVM	BS	C5.0
1	56.5217%	56.5217%	52.1739%	65.2174%
2	47.8261%	56.5217%	56.5217%	47.8261%
3	43.4783%	65.2174%	56.5217%	52.1739%
Item #	Cluster 2			
	NN	SVM	BS	C5.0
1	62.5000%	56.2500%	50.0000%	68.7500%
2	62.5000%	62.5000%	50.0000%	43.7500%
3	50.0000%	50.0000%	31.2500%	56.2500%
Item #	Cluster 3			
	NN	SVM	BS	C5.0
1	48.1481%	51.8519%	48.1481%	48.1481%
2	59.2593%	59.2593%	55.5556%	55.5556%
3	55.5556%	55.5556%	62.9630%	59.2593%
Item #	Cluster 4			
	NN	SVM	BS	C5.0
1	78.2609%	60.8696%	78.2609%	69.5652%
2	43.4783%	43.4783%	43.4783%	43.4783%
3	60.8696%	60.8696%	60.8696%	52.1739%
Item #	Cluster 5			
	NN	SVM	BS	C5.0
1	54.5455%	54.5455%	54.5455%	72.7273%
2	45.4545%	54.5455%	36.3636%	54.5455%
3	63.6364%	63.6364%	45.4545%	63.6364%

Table 3.9(c) Dataset C

Item #	unclustered			
	NN	SVM	BS	C5.0
1	60.0000%	56.0000%	60.0000%	60.0000%
2	55.0000%	53.0000%	59.0000%	63.0000%
3	54.0000%	55.0000%	55.0000%	63.0000%
Item #	Cluster 1			
	NN	SVM	BS	C5.0
1	53.8462%	53.8462%	61.5385%	61.5385%
2	61.5385%	46.1538%	53.8462%	76.9231%
3	46.1538%	69.2308%	53.8462%	76.9231%
Item #	Cluster 2			
	NN	SVM	BS	C5.0
1	68.7500%	56.2500%	75.0000%	43.7500%
2	50.0000%	50.0000%	62.5000%	31.2500%
3	43.7500%	37.5000%	43.7500%	50.0000%
Item #	Cluster 3			
	NN	SVM	BS	C5.0
1	54.8387%	54.8387%	61.2903%	61.2903%
2	48.3871%	67.7419%	67.7419%	67.7419%
3	35.4839%	35.4839%	48.3871%	38.7097%
Item #	Cluster 4			
	NN	SVM	BS	C5.0
1	73.0769%	69.2308%	73.0769%	73.0769%
2	65.3846%	65.3846%	57.6923%	65.3846%
3	76.9231%	73.0769%	69.2308%	73.0769%
Item #	Cluster 5			
	NN	SVM	BS	C5.0
1	42.8571%	42.8571%	50.0000%	50.0000%
2	21.4286%	21.4286%	35.7143%	57.1429%
3	28.5714%	35.7143%	35.7143%	57.1429%

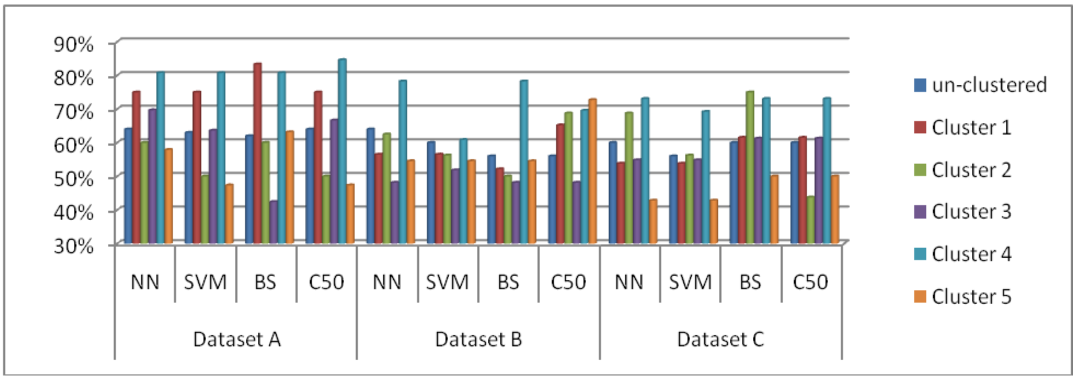

Figure 3.5. Accuracy rate for the first item compared to each dataset and each classification technique.

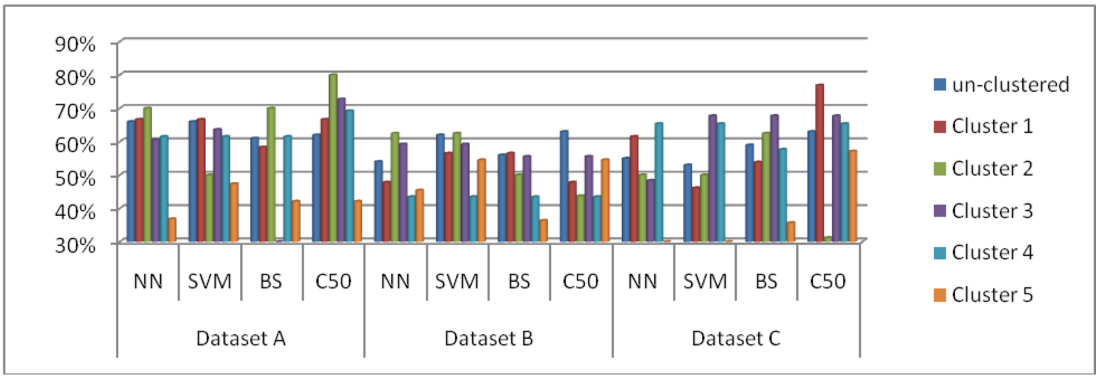

Figure 3.6. Accuracy rate for the second item compared to each dataset and each classification technique.

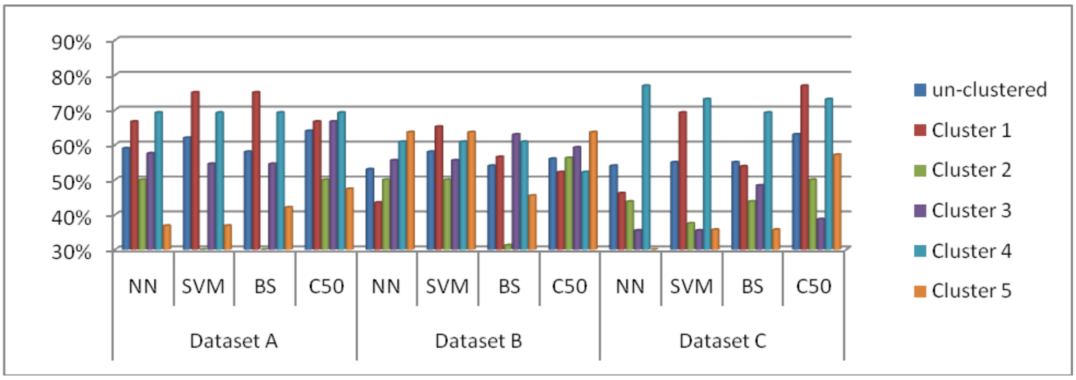

Figure 3.7. Accuracy rate for the third item compared to each dataset and each classification technique.

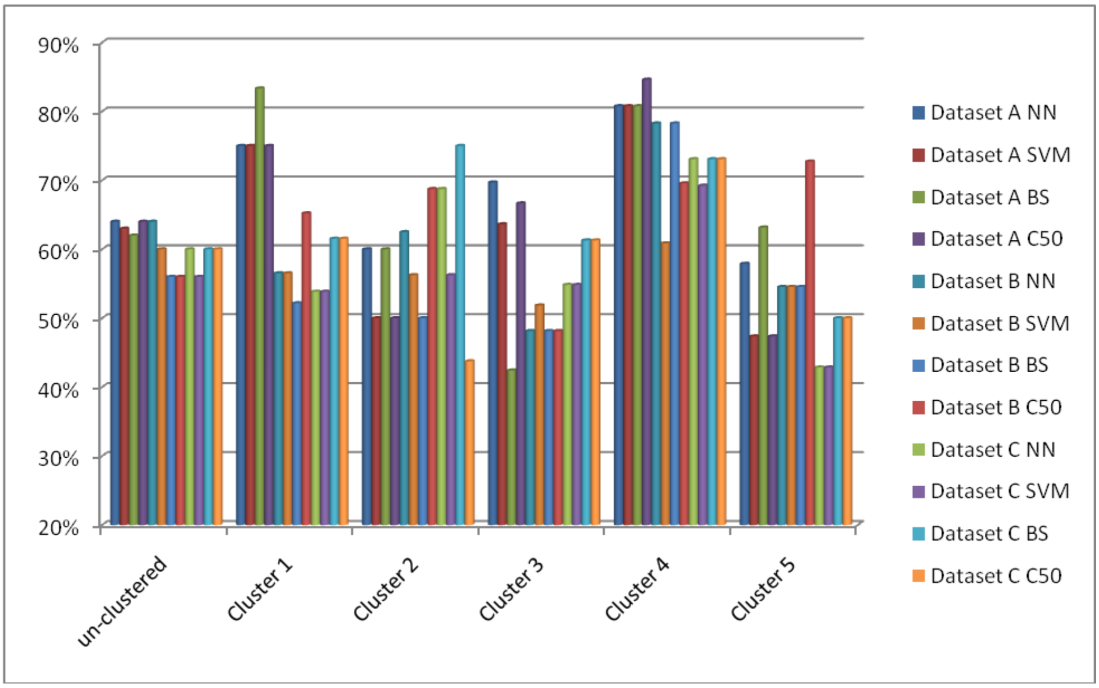

Figure 3.8(a). Accuracy rate for the first item compared to each cluster for all datasets.

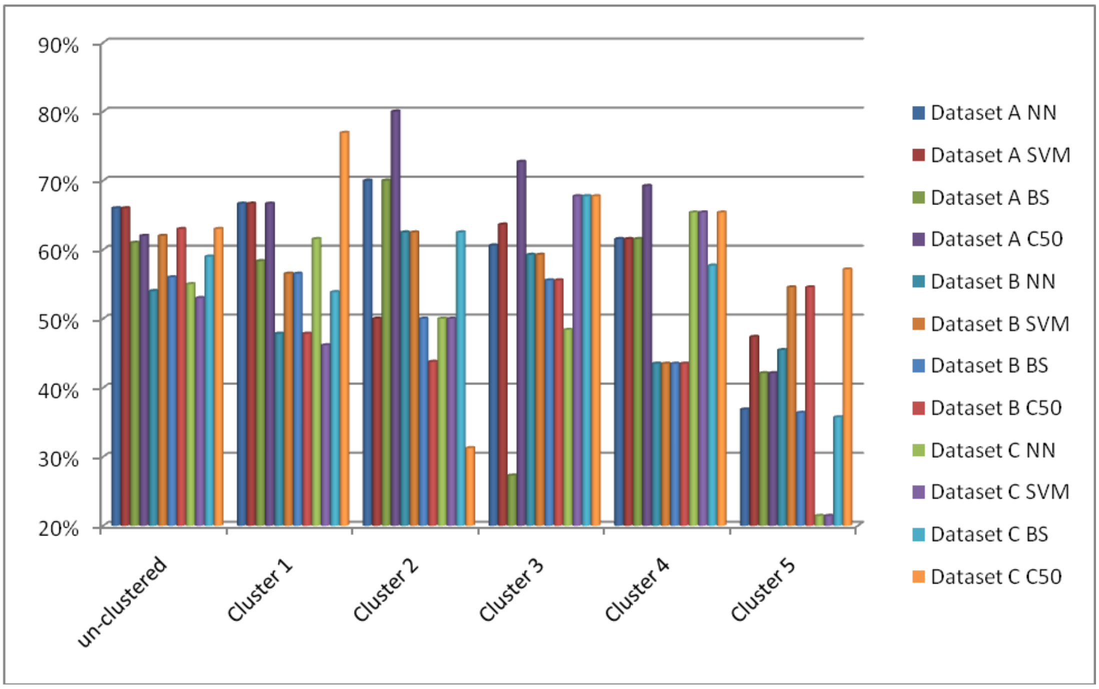

Figure 3.8(b). Accuracy rate for the second item compared to each cluster for all datasets.

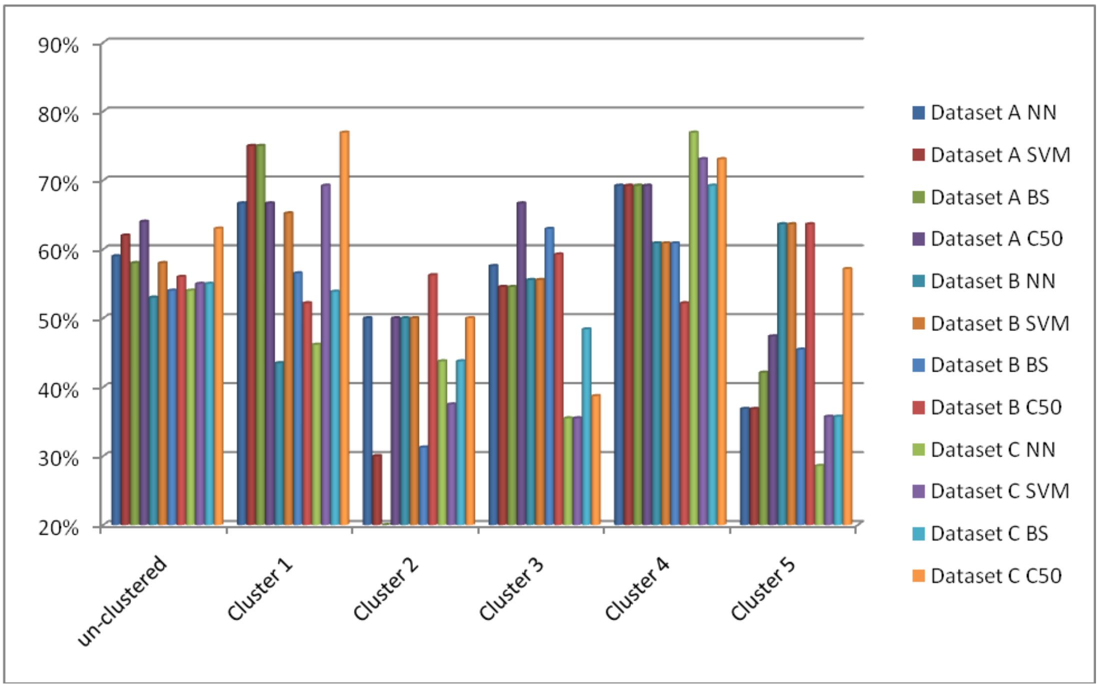

Figure 3.8(c). **A**ccuracy rate for the third item compared to each cluster for all datasets.

3.3.1.3 Classification Model Selection for Mobile Content Filtering

From the classification results, it can be seen that each dataset and each cluster showed the different results inconsistently. As a result, it cannot be concluded which model was the most suitable for mobile content recommendation for the top items. The classification models were varied from data, variables and conditions.

Therefore, the measurement of each cluster and each classification model for each dataset was needed for justification in terms of model selection. The CMScore, the Classification Model Score, was built in this experiment in order to generalise the results and choose the appropriate classification model. The purpose of this measurement was to find which classifier was the most suitable technique used for a cluster-based mobile content recommendation model. This is essential as different datasets and characteristics may require the use of different types of classification techniques. The CMScore can be applied with different number of items, because it concerns accuracy rate derived from predicted items by each classification technique.

The scores for each classifier were calculated based on the accuracy rate results and the ranking of the items which were first, second and third in each dataset. In this metric, the weight for ranking the items was denoted as 3 points for the first ranked item, 2 points for the second ranked item and 1 point for the third ranked item respectively. The number of cases in each cluster was also weighted in the metric for generalisation of the score based on clustering. The CMScore is as shown below:

$$CMScore = \sum_{i=1}^{c} \frac{CS_c \bullet AC_c}{n}$$

where

CS = number of cases in cluster c

AC = accuracy rate of cluster c

c = cluster number

n = number of items.

After that, the CMScore for each classification technique was derived, and then the model selection could use these scores to justify what was the appropriate technique to be used for predicting the top items in mobile content recommendation.

CMScore is used mainly to classify users into the correct mobile content user groups using appropriate technique. Although there may be some variation between various numbers of cases in each class using different techniques, CMScore will try to generalise the score. The main purpose is to classify new users or first time users in the group who have similar preferences. The CMScore results are shown as below in Table 3.10.

Table 3.10

CMScore and Weight Ranking Items

	NN	SVM	BS	C5.0
CMScore	56.4540	56.5894	54.7105	60.8865
Weight ranking	116.6259	114.3763	112.801	123.4508

The CMScore showed that the highest score for the different classification techniques in this dataset was Decision Tree with C5.0. In the second stage of the CMScore measurement, the ranking factor was used by adding the weight of 3 for the first ranked item, 2 for the second ranked item and 1 for the third ranked item. The score again showed that Decision Tree has a higher score than any other technique based on the CMScore weighted ranking as shown in Figure 3.9.

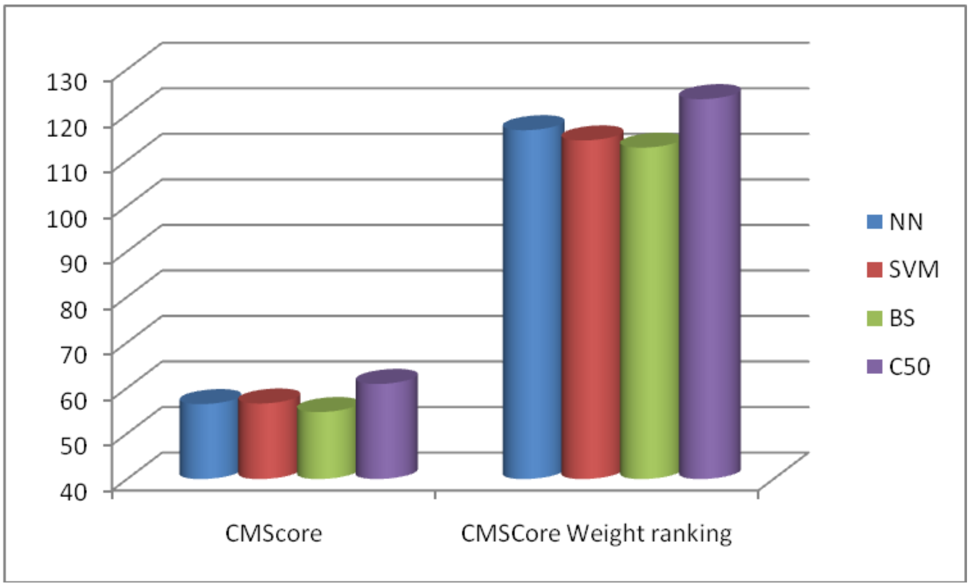

Figure 3.9. The CMScore and CMScore with weight ranking.

3.3.2 Mobile Content Filtering for Finding Relevant Items on the Recommendation System

To complete the process of establishing an integrated model for a mobile content recommendation system, after the top three content items are derived from classification techniques, the next phase is focused on finding relevant items to fulfil a mobile content recommendation system. The technique to address this issue will be investigated in this section.

In data mining, association rules which was proposed by Agrawal et al. [101] has been widely used to find the relationship among product items on the databases. Similar to content items for the mobile Internet, it can be seen as product items, and association rules can be implemented to find the content items.

The objective in using association rules for this stage is to fill the remaining top items with un-rated or new items for mobile content recommendation. This will also allow some relevant low ranking items to be presented to the user too. This is an important stage, as most mobile recommendation systems do not focus on this part. When implementing the association rule technique in the system, if the original association rule would be applied, it may take time to extract the rules, because it will find all possible rules from content items in the database. In addition, it will generate too many rules for the purpose of this research.

In this research, adaptive or alternative association rules should be applied for more efficient rule extraction and to produce appropriate number of rules. The solution for this problem can be partitioning and targeting. Partitioning can help to reduce the number of itemsets to be counted rather than those items in the entire database [79]. The other alternative to increase association rule efficiency is the targeting approach. The association rule can be generated to satisfy measures such as support and confidence with identifying target items. The Classification Association Rule (CAR) [78, 79] is also an alternative for this approach. However, this technique can be used for solving classification problems in a known-class database. In order to find desired items for mobile content recommendation based on association rules, an alternative approach can help in the rule extraction.

The multi-level association rules [79] is an adaptive association rule technique which works by dividing the level for extracting the rules. It uses the hierarchical concept where the higher levels of frequent itemsets have more support than the lower levels. The minimum support in the same level is identical. The advantage of this method is that no complete rule processing is required, as the frequent itemsets in the higher levels help to filter the itemsets in the lower levels with less minimum support. This saved significant computational time in extracting the association rules. Figure 3.10 shows an overview of the rules generator module on an integrated mobile content filtering model. The AR generator obtains input from the previous components which are the user's group identification and top content items filtering. This input includes cluster information, users' rating of content items and predicted top content items. Then, the association rules are extracted and consolidated to find the set of rules for mobile content filtering. The detail of rule consolidation will be shown in

Figure 3.11. After that, these rules will be used to find the relevant content items for a mobile content recommendation generator.

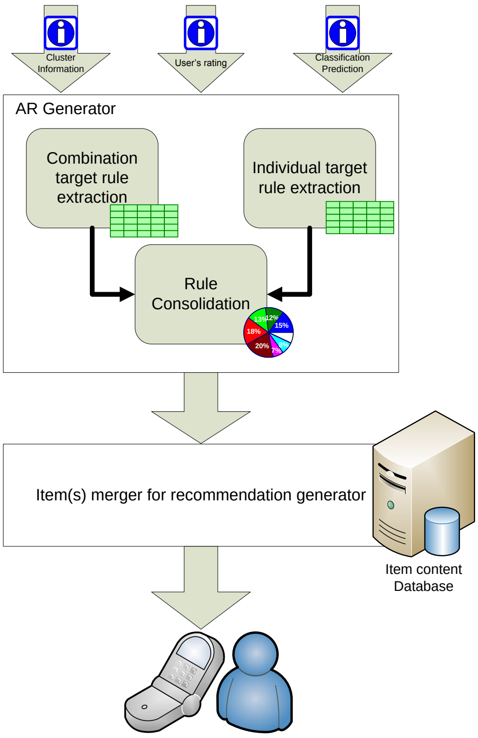

Figure 3.10. Extracting of relevant items module based on the association rules process.

3.3.2.1 The Proposed Multi-level Targeting Classification Association Rule Technique (MTCAR)

With the discussion in the last section, one can see that there is no one association rules technique that can be used to solve all the problems required in the mobile recommendation system. The proposed methodology is to find the association rules on the relevant content items for the mobile content filtering in the recommendation system by combining the classification association rule and multi-level association rules. The purpose is to reduce the number of redundant rules and classify relevant content items based on classification and clustering techniques. With all this, the proposed system will address the current limitation of mobile content recommendation in the early stage. It also enhances the system by finding the relevant items for the user based on user profile.

According to the experiment dataset, there is no information related to cluster and class. However, in the previous phase as described earlier, mobile content user analysis using clustering has been done to find the group based on similar demographic factors. In addition, a classification technique has been incorporated to predict the top most wanted items based on cluster information. Then, from the classification results, these can be used as targets and antecedents to find the association rules from datasets.

In the proposed multi-level approach, the first level will deal with the top ranking items. This stage implements the concept of classification association rules to find the relevant items that are related to the top ranking items. With the top ranking items derived from the classification phase, they are defined as the targets in the rule extraction process. In the first level, only the top three ranked items are used as the target. In the second and subsequent levels of association rules, the rules for the level are extracted by setting target from the top level, which is the precedent top ranking items.

After the rules for the different levels have been extracted, the next step is rule consolidation. The first step is using rules from the first level to find the target items based on top N. If the system can find relevant items up to top N, it is stopped. In contrast, if the first level rule cannot complete the requirement the system goes to the next level and finds the target according to the ranking of the content item in each cluster, specifically, first, second and third. In addition, if the rule and target are duplicated from the first level, it would be cut off. Finally, the recommended items are derived and prepared to be pushed into a mobile recommendation system. Figure 3.11 shows the rule consolidation process.

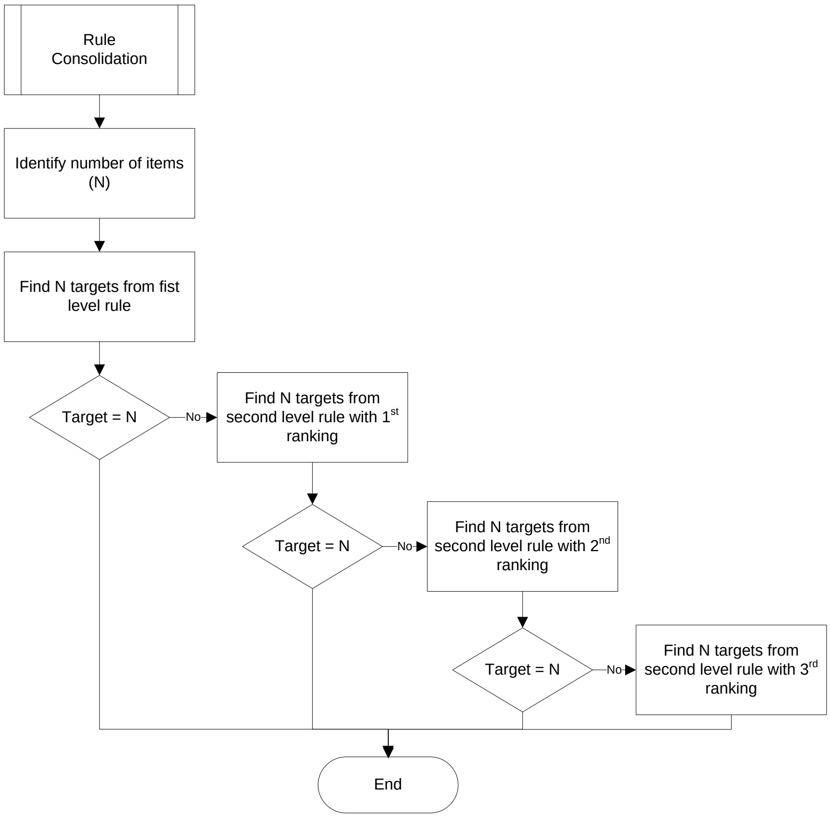

Figure 3.11. Rule consolidation process.

3.3.2.2 The Methodology and Experiment for Proposed MTCAR

To construct the association rules for relevant items on a recommendation system for a first time user establishing a model, the same datasets have been used as those used in the clustering and classification experiments. The top three items based on a user's ranking will be used as targets of classification association rules for the first step. In the datasets, it has been transformed as a binary value (0, 1) to identify the contents that a user prefers. This is similar to the classification phase. The aim of this phase is to recommend relevant items that may not have been rated highly but might still be of interest to the users based on their profile.

The Apriori algorithm has been implemented in this phase. For the first step of MTCAR, the minimum support and confidence are set at 50% in each level respectively. Then, the rules for the first level with three antecedents are gained. After that, the second level will be run separately for each target item based on its ranking, specifically, first, second and third. With a lesser minimum support and confidence, the results of the second level are the rules from each item.

So, from the first level and the second level, all rules will be consolidated to rank the outcome sorting by level (first or second) and order (for the second level). The duplicate rules are eliminated and the rules that show the same result are also cut off using support, confidence, level and sequence.

3.3.2.3 Results of the MTCAR

This section presented the results of extracting association rules to find the relevant items for the mobile recommendation system. Table 3.11 showed an example of the extracted rules from the first level MTCAR. 'Consequent' represented the relevant items which were derived from target items in the 'Antecedent' regarding the rules. Notice that 'IM' stands for item number.

Table 3.11

Example of Association Rule Extraction for Cluster 5 by First-level Rules with Three Content Items

Consequent	Antecedent	Support %	Confidence %
IM312 = 1.0	IM31 = 1.0 and IM311 = 1.0	50.6849	97.2973
IM312 = 1.0	IM14 = 1.0 and IM31 = 1.0	34.2466	96.0000
IM312 = 1.0	IM14 = 1.0 and IM31 = 1.0 and IM311 = 1.0	34.2466	96.0000
IM312 = 1.0	IM311 = 1.0	54.7945	95.0000
IM312 = 1.0	IM31 = 1.0	53.4247	94.8718
IM312 = 1.0	IM14 = 1.0 and IM311 = 1.0	36.9863	92.5926
IM313 = 1.0	IM14 = 1.0 and IM31 = 1.0	34.2466	84.0000
IM313 = 1.0	IM14 = 1.0 and IM31 = 1.0 and IM311 = 1.0	34.2466	84.0000
IM313 = 1.0	IM31 = 1.0 and IM311 = 1.0	50.6849	83.7838
IM313 = 1.0	IM311 = 1.0	54.7945	80.0000
IM313 = 1.0	IM31 = 1.0	53.4247	79.4872
IM313 = 1.0	IM14 = 1.0 and IM311 = 1.0	36.9863	77.7778

Then, the second level of rule extraction was shown from Tables 3.12 to 3.14 with the ranking item from first level targets. They were first, second and third ranking respectively. Please note that IM stands for item number.

Table 3.12

Example of Association Rule Extraction for Cluster 5 by Second Level Rules with First Ranking Item

Consequent	Antecedent	Support %	Confidence %
IM311 = 1.0	IM14 = 1.0	52.0548	71.0526
IM31 = 1.0	IM14 = 1.0	52.0548	65.7895
IM312 = 1.0	IM14 = 1.0	52.0548	65.7895
IM313 = 1.0	IM14 = 1.0	52.0548	55.2632
IM23 = 1.0	IM14 = 1.0	52.0548	52.6316
IM32 = 1.0	IM14 = 1.0	52.0548	52.6316
IM326 = 1.0	IM14 = 1.0	52.0548	52.6316
IM11 = 1.0	IM14 = 1.0	52.0548	50.0000

Table 3.13

Example of Association Rule Extraction for Cluster 5 by Second Level Rules with Second Ranking Item

Consequent	Antecedent	Support %	Confidence %
IM312 = 1.0	IM311 = 1.0	54.7945	95.0000
IM31 = 1.0	IM311 = 1.0	54.7945	92.5000
IM313 = 1.0	IM311 = 1.0	54.7945	80.0000
IM14 = 1.0	IM311 = 1.0	54.7945	67.5000
IM326 = 1.0	IM311 = 1.0	54.7945	60.0000
IM315 = 1.0	IM311 = 1.0	54.7945	55.0000
IM32 = 1.0	IM311 = 1.0	54.7945	55.0000
IM316 = 1.0	IM311 = 1.0	54.7945	50.0000
IM327 = 1.0	IM311 = 1.0	54.7945	50.0000

Table 3.14

Example of Association Rule Extraction for Cluster 5 by Second Level Rules with Third Ranking Item

Consequent	Antecedent	Support %	Confidence %
IM312 = 1.0	IM31 = 1.0	53.4247	94.8718
IM311 = 1.0	IM31 = 1.0	53.4247	94.8718
IM313 = 1.0	IM31 = 1.0	53.4247	79.4872
IM326 = 1.0	IM31 = 1.0	53.4247	64.1026
IM14 = 1.0	IM31 = 1.0	53.4247	64.1026
IM315 = 1.0	IM31 = 1.0	53.4247	53.8462
IM32 = 1.0	IM31 = 1.0	53.4247	53.8462
IM327 = 1.0	IM31 = 1.0	53.4247	51.2821

Finally, the result of the rule consolidation between the first level and the second level of rules were shown in Table 3.15. The results were the content items with the top ten ranking. The top three were the results generated from the classification phase and the rest are generated by MTCAR in recommending relevant items. These results would be shown in the recommendation system.

Table 3.15

Example of the Result of Rule Consolidation for Cluster 5 by MTCAR

Rank	Results
1	IM14
2	IM311
3	IM31
4	IM312
5	IM313
6	IM23
7	IM32
8	IM326
9	IM11
10	IM315

To prove how the integrated model of mobile content recommendation works, the model verification is needed and it will be discussed in Chapter 4.

3.4 Discussion and Conclusion on the Integrated Mobile Content Recommendation Model

Access through the mobile Internet with content filtering is the one feature that a mobile device user needs to be served. Exchanging some allow-to-provide data, such as fundamental information like demographic factors, can help a recommendation system provide personalised content, which has been filtered from available data. This can serve new users and help them experience mobile personalised content by a recommendation system.

In this chapter, the knowledge of mobile content filtering for a recommendation system has been extracted to address the problem of a new user connection or the first time user. Based on primary data, it can be represented in terms of statistics and show the user information, including user rating with ordinal. This can establish the integrated model for this kind of user for a non-interactive recommendation system with multiple processes.

The model has brought the advantages of clustering techniques to group the mobile content user for a user-based system. Clustering analysis can help the mobile content recommendation system classify a user group based on criteria, and demographic factors are the important criteria that can be used to determine a user group. Thus, each cluster is identified to be a group of a new user or an active user in the recommendation system. To proceed with clustering techniques, partition clustering is the most common to apply on cluster analysis but the problem is the number of cluster determinations. Zoning-Centroid can help to determine number of clusters to facilitate cluster analysis for mobile content recommendation by separation distance between the cluster centre and data into zones in order to form the cluster. This can help unsupervised learning data to identify the class.

After the mobile content user analysis using clustering, the next phase to be done is mobile content filtering. The first step for this phase is finding the top content items for mobile content filtering in each cluster using classification techniques. This can help mobile content recommendation to predict the target items for the item-based concept. In this stage, different classification techniques have been compared. The experiment showed the different classifiers that represent inconsistent results. The factors that influence the inconsistency in a mobile content classification model are data characteristics and data type. Because this data source came from a survey of user preferences primarily, the variety of values in variables, such as personal interest and preferences, is high and varied from person to person. Furthermore, this data has been transformed to categorical data to analyse in a clustering classification process. As a result, CMScore which is the measurement to find the appropriate classifier needs to be built. The results have shown that Decision Tree with C5.0 is the best, and most consistent, model to predict the top items for a mobile content user.

The second step for the mobile content filtering phase is finding relevant items to enhance a mobile content recommendation system. Association rules help to find relevant items by extracting rules based on support and confidence. The higher the level of confidence, the more confidence will be correlated between content items.

However, it produces an enormous and impractical number of rules. The proposed MTCAR facilitates rule generation with level separation and target determination by top items from the classification of the top content items phase. In MTCAR, it provides two levels of association rules by identifying multiple targets for high support and confidence in the first level. Then, the second level provides rules to support consequent based on the number of top N that the recommendation needs. This is to find the relevant content items for a recommendation system. Numbers of rules are reduced significantly compared to traditional association rules, and redundant rules are cut off by the rule consolidation process. In addition, rule complexity is reduced by level dividing as well. This can enhance the mobile content recommendation system.

Finally, in this chapter establishing the integrated model to fulfil prediction and recommendation for a mobile content recommendation system has been proposed. This can address the problems of the first rater and the new user connection for a non-interactive recommendation system. The proposed methodology has covered recommendation processes which are prediction and recommendation. In addition, it shows the process that concerns the first time user connection, and the method not only predicts top content items but also retrieves relevant items. These enhance the mobile content recommendation system. The next chapter will discuss the verification of the model being established.

CHAPTER 4

INTEGRATED MOBILE CONTENT RECOMMENDATION: A COMPARISON STUDY

4.1 Introduction

After establishing the integrated mobile content recommendation in the last chapter, the next phase is to perform a comparison study with some common techniques used for mobile content recommendation found in the literature. As the integrated mobile content recommendation system also takes care of the first rater item and the new user of the system, a comparison study to include these would be provided. The results from the previous chapter will be used again in this chapter, which include the mobile user identification, prediction of top ranking items, and finding the relevant items. All the results from each stage are combined as the final suite of methods in the integrated mobile recommendation system. The metrics used to measure the performance of each method will be compared as well. At the end of this chapter the discussion of the comparison study will be shown.

4.2 Comparison Study

This section presents methodology and metrics for recommendation measurement.

4.2.1 Methodology and Comparison Techniques

The first stage of the methodology begins with the pre-processing of the data such that the data will be suitable for each comparison technique. The datasets chosen are a duplication of the datasets used in the previous experiment that was performed for the MTCAR method. They are Dataset A, Dataset B and Dataset C. They are randomised from 400 records to form the dataset. Three hundred records will be used as a training set and 100 records will be used as a testing set. Although the three datasets are, as previously stated, similar to those used in MTCAR, they contain some small formatting differences. The pre-processing phase is required to format the data to ensure suitability for each comparing technique. The second stage compares the integrated method with other techniques by using the same dataset for both training and testing in each method. The techniques to be compared with are collaborative filtering and association rules. The details of these techniques will be described in the following:

Collaborative Filtering

The most widely used technique in the recommendation system is collaborative filtering. The item-based collaborative filtering is chosen because the research of Papagelis and Plexousakis [60] has shown that the item-based algorithm performed better than the user-based algorithm. Moreover, the study of Yu et al. [54] presented that using the Pearson Correlation Coefficient performed better than the Kendall Correlation, and positive correlated neighbours gain higher accuracy.

Firstly, the Pearson Correlation Coefficient is used to analyse the similarity among the items available on the training dataset. The similarity matrix of items is derived. Secondly, in the recommendation stage, an analysis of finding the highest similarity measure will be performed, and this pushes items with a higher measure to the recommendation list. The recommendation list is derived by eliminating duplicated items; a duplicated item with a lower similarity measure is removed. The recommendation list is then compared with testing data. In addition, the testing data is separated from training data. This is similar to other comparing methods in the experiment. The technique described above is a memory-based recommendation system.

Association Rules

This technique is a model-based approach for a recommendation system. It is an appropriate technique used to find associated items or relevant items for the system. This technique constructs the rules and produces the consequences, that is, the results of the relevant items according to antecedents or the conditions of the mobile content. The association rules with Apriori algorithm are constructed using the same criteria as MTCAR, as is the support and confidence level. The rules are constructed from the three training datasets. The association rules will be used with the testing datasets by demographic factors input. After that, the recommendation system lists are derived from the consequent of rules according to the user profile.

4.2.2 Metrics for Recommendation Measurement

The most well-known metric to measure a recommendation system is accuracy rate [10]. This metric is widely used to identify the quality of the model or system. Generally, the formula for this metric can be formed as follows:

$$accuracy = \frac{number\_of\_correction}{total\_case}.$$

In the Olmo and Gaudioso study [10], they reformulate accuracy metric by considering that p and P are binary functions. Where P refers to prediction of recommendation system and p refers to the real preferences. Both functions will offer values 1 which refers to a successful recommended item and 0 for any other case. This is the metric for a recommendation system:

$$accuracy = \frac{\sum_{(\forall u,i / r(u,i)=1)} 1 - \left| p_{(u,i)} - P_{(u,i)} \right|}{R}$$

The other metric that is commonly used for recommendation is Mean Absolute Error (MAE):

$$MAE = \frac{\sum_{u,i} \left| p(u,i) - P(u,i) \right|}{N}$$

where *p(u,i)* is predicted value and *P(u,i)* is the actual value of user rating regarding an item. In this metric, if the MAE is low, it means the performance of the recommendation system would be better than the higher MAE. There are many researchers using MAE as a metric to measure performance of the recommendation system [54, 56, 60, 65, 104].

4.3 Experimental Results

This section shows experimental results of recommendation system performance and MTCAR performance comparison with association rules generation. After that, the qualitative comparison of MTCAR is presented.

4.3.1 Recommendation System Performance

The comparison results are shown in Tables 4.1(a) to (c) and the graphs in Figures 4.1(a) to (c). They represent the accuracy rate of the recommendation system for the top 10 items in each dataset.

Table 4.1(a)

Accuracy Rate of Dataset A

Techniques	**n=3**	**n=4**	**n=5**	**n=6**	**n=7**	**n=8**	**n=9**	**n=10**
MTCAR	64.67%	62.75%	57.60%	54.50%	54.86%	53.63%	52.67%	51.30%
AR	64.33%	62.75%	61.40%	57.50%	54.57%	50.75%	47.22%	44.70%
CF	61.67%	56.75%	52.20%	51.00%	48.43%	44.75%	43.33%	42.70%

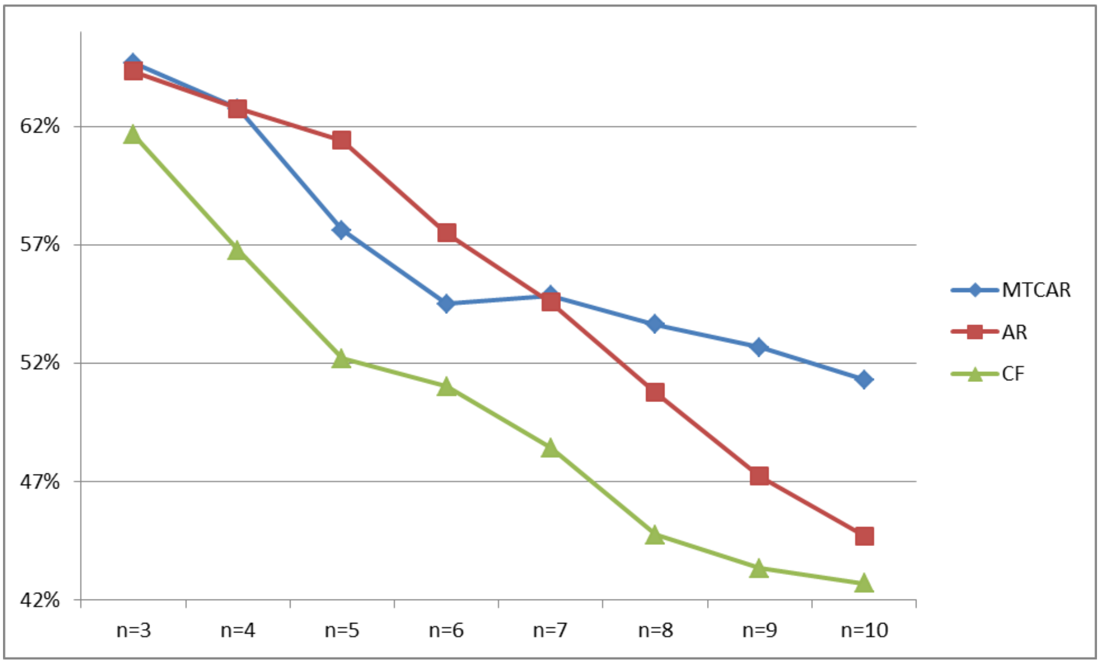

Figure 4.1(a). A comparison of the accuracy rate between MTCAR and other techniques in Dataset A.

Table 4.1(b)

Accuracy Rate of Dataset B

Techniques	n=3	n=4	n=5	n=6	n=7	n=8	n=9	n=10
MTCAR	56.00%	53.50%	51.60%	49.00%	47.57%	46.75%	46.11%	44.50%
AR	56.33%	54.25%	53.40%	50.83%	48.00%	45.00%	42.67%	39.90%
CF	54.33%	52.50%	49.80%	46.67%	44.14%	41.50%	39.44%	39.50%

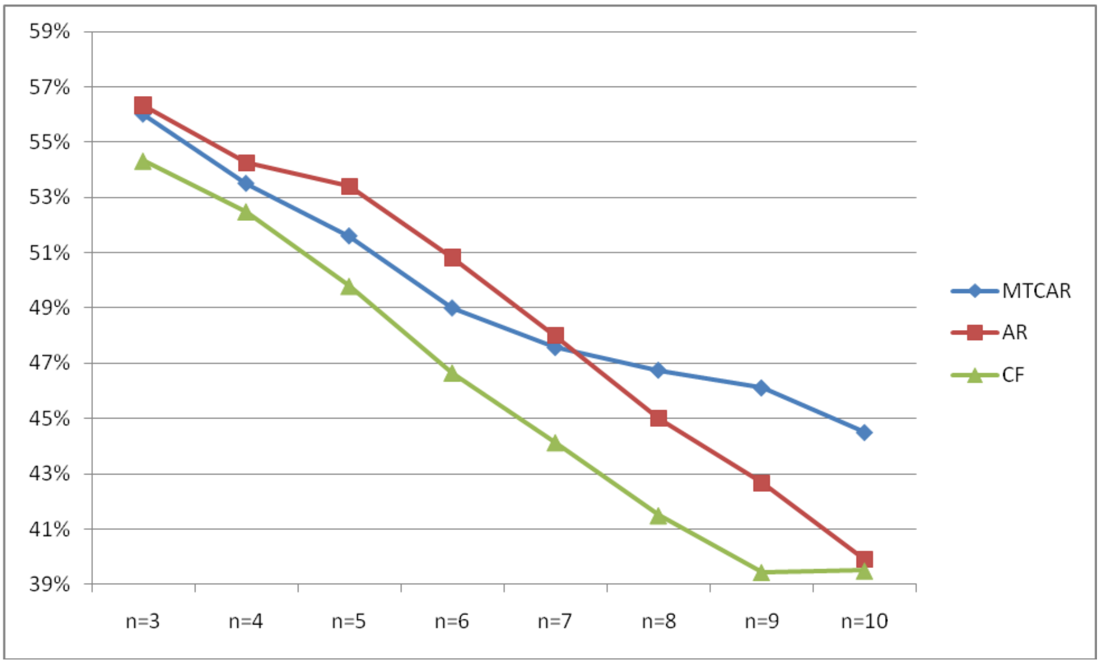

Figure 4.1(b). A comparison of the accuracy rate between MTCAR and other techniques in Dataset B.

Table 4.1(c)

Accuracy Rate of Dataset C

Techniques	**n=3**	**n=4**	**n=5**	**n=6**	**n=7**	**n=8**	**n=9**	**n=10**
MTCAR	60.33%	58.50%	55.80%	53.67%	51.71%	50.38%	48.67%	47.30%
AR	57.33%	57.25%	56.40%	52.33%	49.71%	47.25%	44.33%	41.30%
CF	61.33%	57.00%	56.20%	53.50%	51.00%	47.63%	45.44%	45.30%

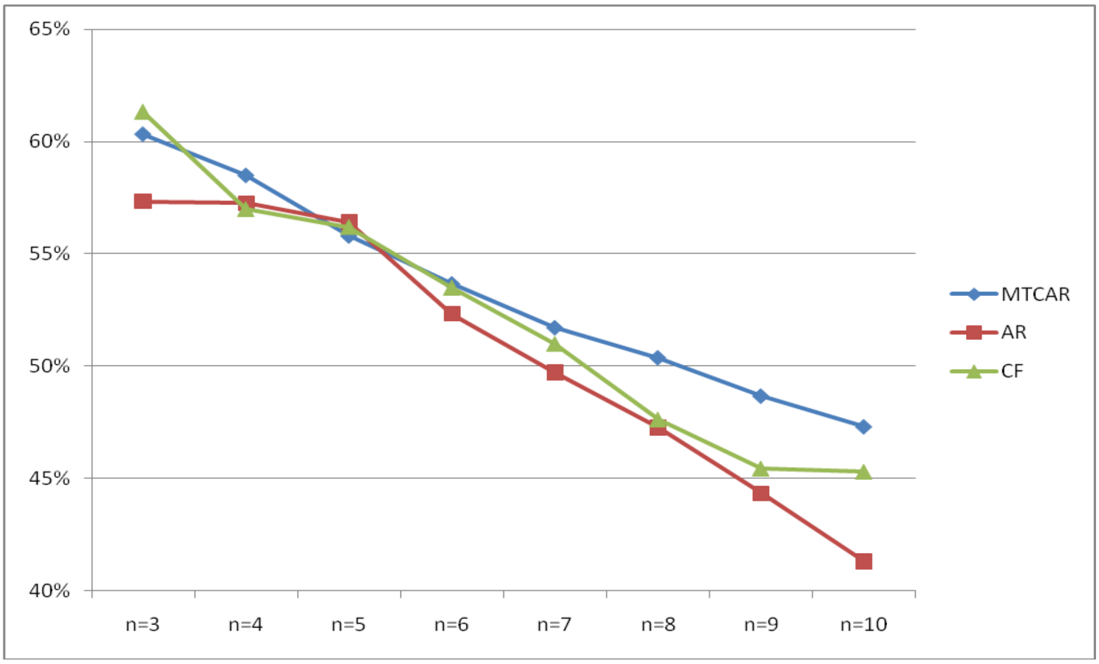

Figure 4.1(c). A comparison of the accuracy rate between MTCAR and other techniques in Dataset C.

The comparison results show that MTCAR outperformed the collaborative filtering technique. When comparing with the association rules, the first three or four items seem to be similar to association rules in Datasets A and B, with better performance in Dataset C. From the top three to the top six items, the three methods have comparable results, but after the top six MTCAR can perform better in finding relevent items. It shows significant results in the accuracy rate.

Having considered the accuracy rate which is the standard benchmark for a recommendation system, it is necessary to consider Mean Absolute Error (MAE) when comparing the system's performance. Figures 4.2(a) to (c) present MAE in each technique for each predicted item.

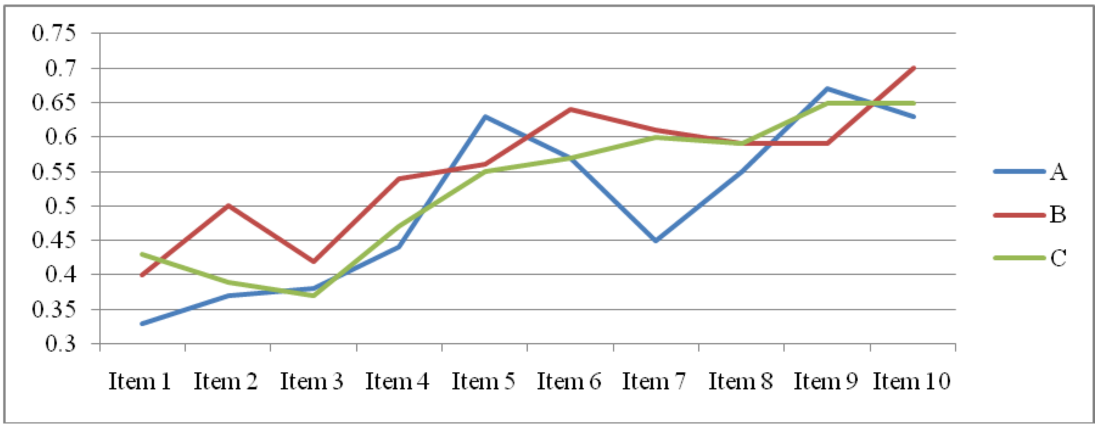

Figure 4.2(a). A comparison of Mean Absolute Error (MAE) between datasets for the MTCAR technique.

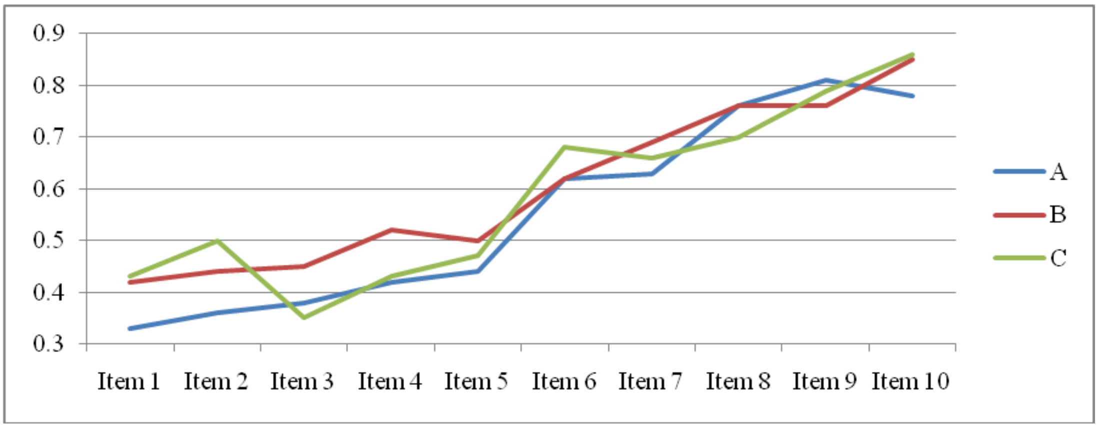

Figure 4.2(b). A comparison of Mean Absolute Error (MAE) between datasets for the association rules technique.

Figure 4.2(c). A comparison of Mean Absolute Error (MAE) between datasets for the collaborative filtering technique.

Generally, the results of MAE in each technique for each dataset presented a similar trend, which was a low MAE from the start and a significant increase when the number of items was higher. Only the collaborative filtering from item's 9 and 10 has a slightly reduced trend. Figure 4.3(a) – Figure 4.3(c) present the cumulative errors in each technique for prediction item from item 3 to item 10.

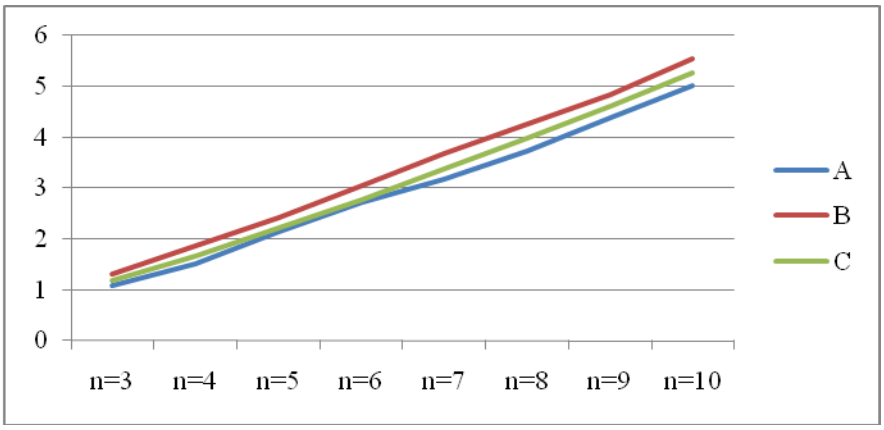

Figure 4.3(a). A comparison of cumulative Mean Absolute Error (MAE) of total prediction items from 3 to 10 for the MTCAR technique.

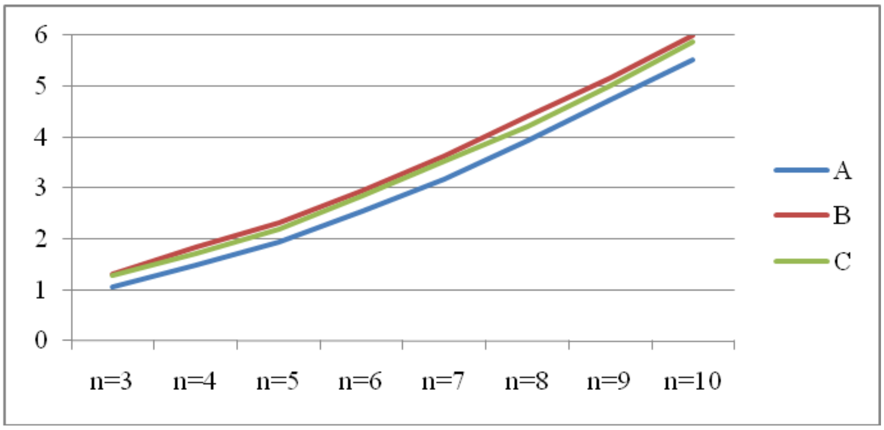

Figure 4.3(b). A comparison of cumulative Mean Absolute Error (MAE) of total prediction items from 3 to 10 for the association rules technique.

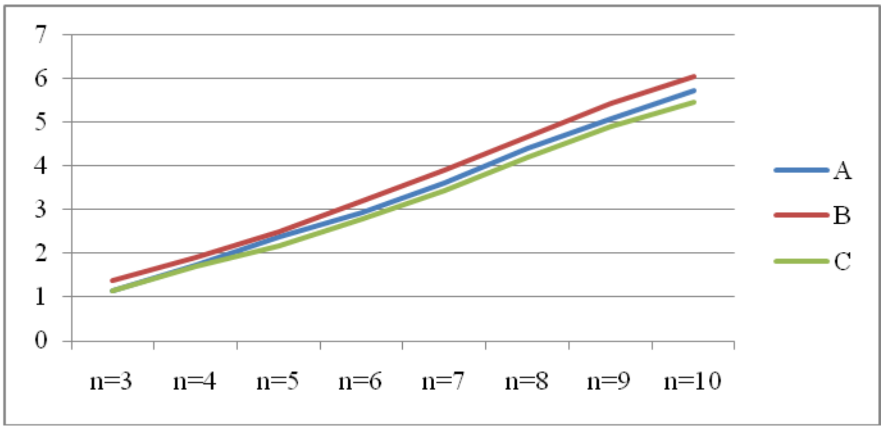

Figure 4.3(c). A comparison of cumulative Mean Absolute Error (MAE) of total prediction items from 3 to 10 for the collaborative filtering technique.

The collective cumulative MAE also shows the same trend in each technique and each dataset. The error was lower at the starting point from the total number of prediction items at 3 and increases dramatically to the number of prediction items at 10.

For comparison between MTCAR and other techniques in Mean Absolute Error measurement, Figures 4.4(a) to (c) also demonstrate these results.

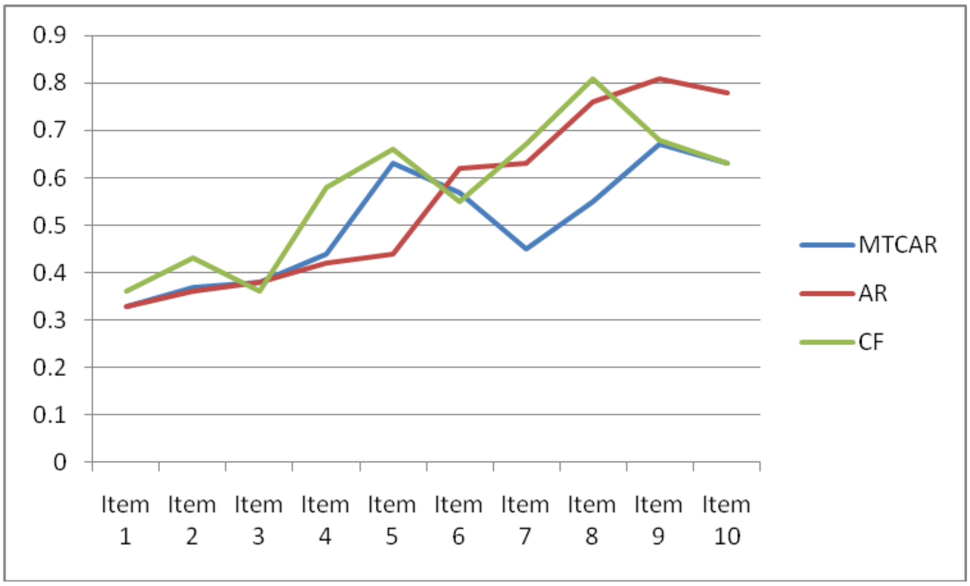

Figure 4.4(a). A comparison of Mean Absolute Error (MAE) between MTCAR and other techniques in Dataset A.

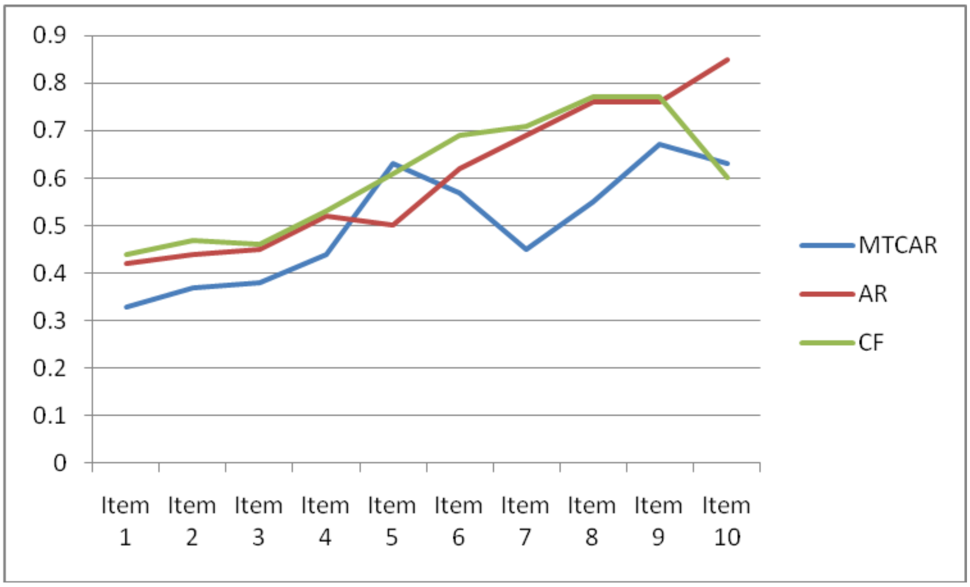

Figure 4.4(b). A comparison of Mean Absolute Error (MAE) between MTCAR and other techniques in Dataset B.

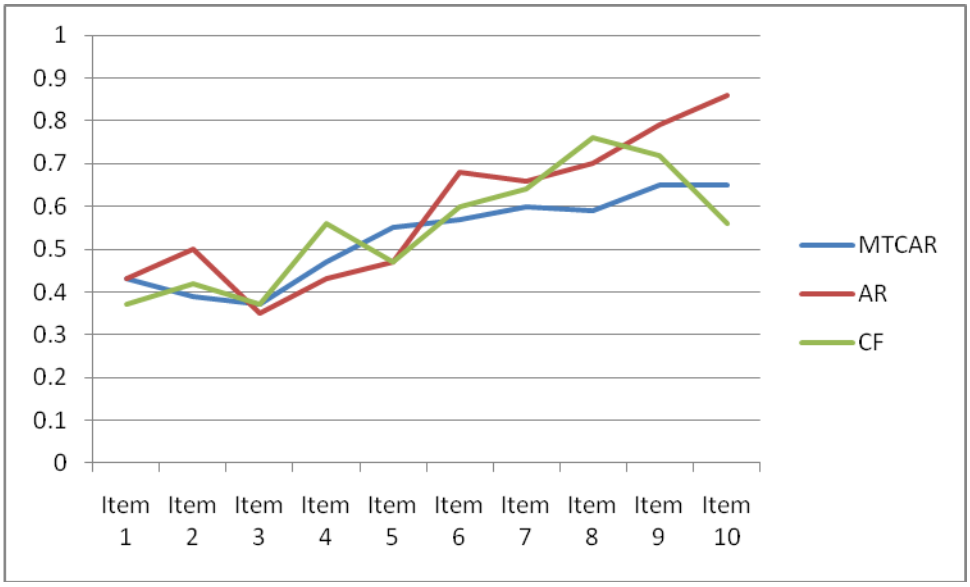

Figure 4.4(c). A comparison of Mean Absolute Error (MAE) between MTCAR and other techniques in Dataset C.

According to the graphs above, it appears that MTCAR mostly returns significantly less individual prediction errors in each item, especially after the sixth item. Only collaborative filtering performed slightly better in the 10th item. Figures 4.5(a) to 4.5(c) show the cumulative MAE of total prediction items compared with other techniques and it can be seen precisely that MTCAR shows less error rate. This means MTCAR can provide better performance of a mobile content recommendation, compared with other techniques.

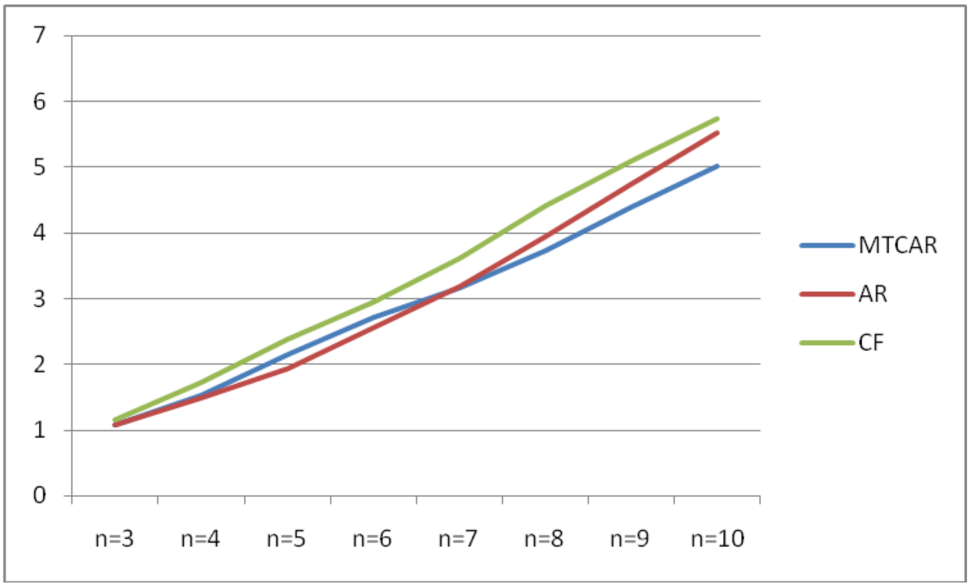

Figure 4.5(a). A comparison of cumulative Mean Absolute Error (MAE) of total prediction items from 3 to 10 between MTCAR and other techniques in Dataset A.

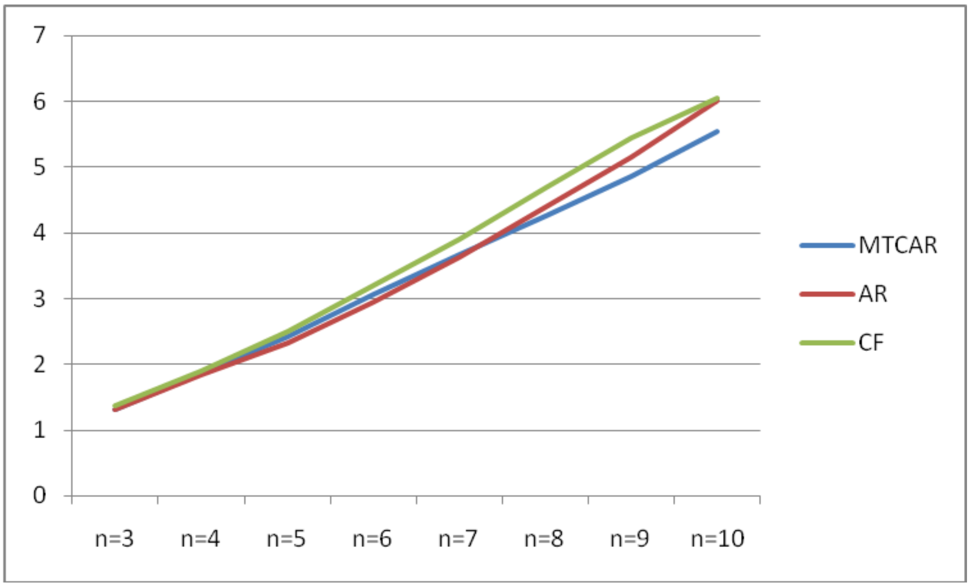

Figure 4.5(b). A comparison of cumulative Mean Absolute Error (MAE) of total prediction items from 3 to 10 between MTCAR and other techniques in Dataset B.

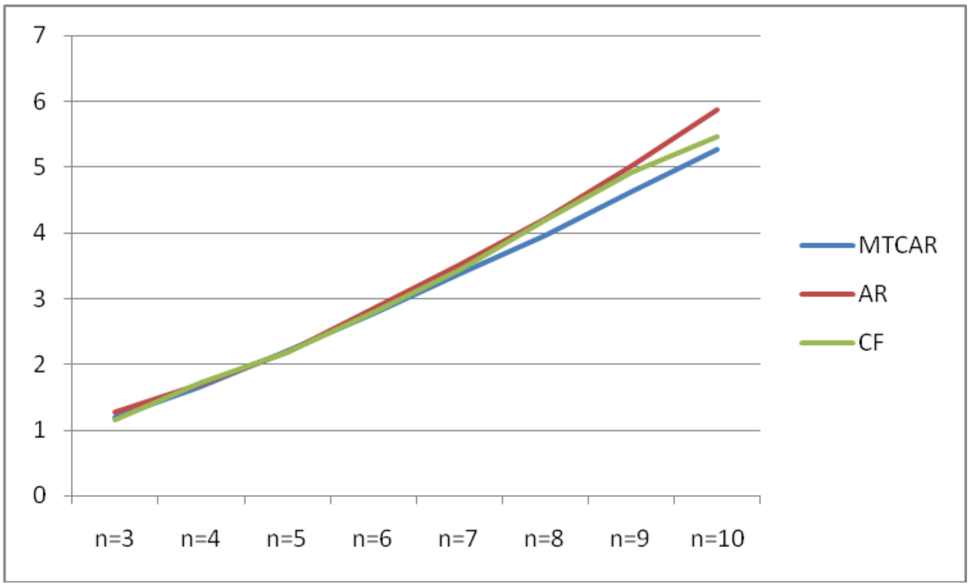

Figure 4.5(c). A comparison of cumulative Mean Absolute Error (MAE) of total prediction items from 3 to 10 between MTCAR and other techniques in Dataset C.

4.3.2 MTCAR Performance Comparison with Association Rules Generation

The other performance of MTCAR compared with association rules can be seen from Table 4.2. It seems that MTCAR generates fewer rules than association rules, and the MTCAR technique provided better performance in a recommendation for mobile content. They are shown on previous results in terms of accuracy rate and Mean Absolute Error. The percentage of rules reduced is calculated from the difference between the number of rules generated by each technique and divided by the number of maximum rules for each dataset.

Table 4.2

Percentage of Number of Reduced Rules Generation

Technique	Dataset A	Dataset B	Dataset C
MTCAR	194	170	190
AR	218	247	211
% of Rules Reduced	11.01%	31.17%	9.95%

Furthermore, Table 4.3 shows that the number of recommended items generated from MTCAR was significantly more compared to the association rules technique. In all datasets, MTCAR gained the number of average recommended items generated at 9.7, 9.38 and 9.5 out of 10 items respectively, whereas association rules obtained an average number of recommended items generated at 7.66, 8 and 7.43 consecutively. Moreover, MTCAR can generate the number of recommended items on a system of around 27% in Dataset A and Dataset C. Although Dataset B showed lower percentage at around 17%, it was still better in terms of the number of recommended items.

Table 4.3

Number of Generated Recommendation Items for Mobile Content

Technique	**Dataset A**	**Dataset B**	**Dataset C**
MTCAR	9.70	9.38	9.50
AR	7.66	8.00	7.43
% of Item Generation	26.63%	17.25%	27.86%

The level of emptiness is shown in Table 4.4. It was the measure to indicate that a recommendation system was unable to generate recommendation items according to available information. It means the system shows ‘empty’ for these items. The measurement was calculated by the sum of empty recommendation items in each user for each dataset as follows.

$$\textit{Percentage of ‘empty’ recommendation} = \frac{\sum_{i=1}^{n} \begin{cases} 1 : Item_{i\,j} = \varnothing \\ 0 \end{cases}}{TotalItem_{n\,j}}$$

where i is the item number of dataset j.

MTCAR can perform better in terms of items generation for a mobile content recommendation system. All datasets show the percentage of emptiness was at 3%, 6% and 5% respectively, and while the association rules showed a much higher percentage compared to the MTCAR technique, the percentage of the difference was shown at 87%, 69% and 81% for each dataset consecutively.

Table 4.4

Level of Emptiness Generation for a Recommendation System

Technique	Dataset A	Dataset B	Dataset C
MTCAR	3%	6%	5%
AR	23%	20%	26%
% of Difference of emptiness	87%	69%	81%

4.3.3 Qualitative Comparison

To verify that MTCAR can be used on a mobile content recommendation system, the qualitative comparison is carried out. The data to compare with the proposed method is collated from a large mobile portal site in Thailand and the statistic has been recorded in www.mobilethai.net. The primary data is 552,898 page views for the mobile portal site and there are various categories, including news, fortune teller and game downloading. The actual ratio of page views is unable to be disclosed; therefore, it can show roughly the proportion. Furthermore, to compare with the data that has been used in this experiment by MTCAR, the page view category will be filtered to find the content that is a match with the experiment data.

As a result, the data on page view was filtered down to news, entertainment, mobile download, and sports. Likewise, the results of the recommendations would be reduced to categories that were similar to the mobilethai.net mobile portal page view for a fair comparison. The results are presented in Table 4.5.

Table 4.5

Proportion of Mobile Content Compared to Actual Mobile Portal Page View

Content	Mobile Portal Page View	Dataset A	Dataset B	Dataset C
News	40.00%	45.75%	45.92%	42.45%
Entertainment	33.33%	25.20%	26.12%	29.89%
Mobile Download	16.00%	19.74%	19.63%	23.21%
Sports	10.67%	9.31%	8.32%	4.45%
Total	100.00%	100.00%	100.00%	100.00%

It seems that the mobile content recommendation from MTCAR reflects the real world for mobile content usage. This can be seen from Figures 4.6(a) to (d). The proportions of the pie charts are quite similar, with each dataset showing the sports content was less compared to the actual proportion on the mobile portal page view, and the mobile download being higher.

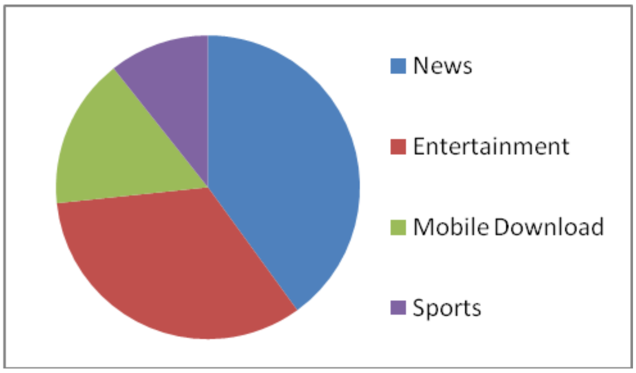

Figure 4.6(a). Proportions of mobile content page view.

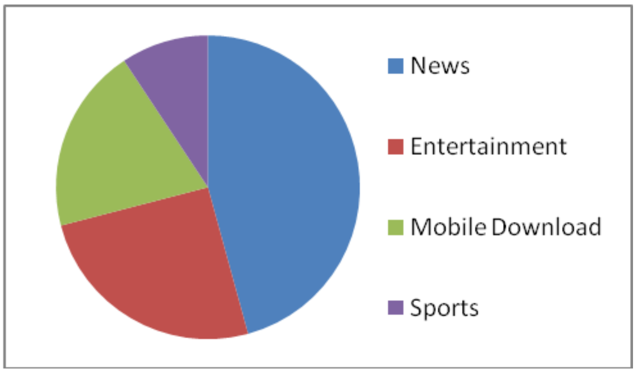

Figure 4.6(b). Proportions of mobile content recommendation from Dataset A.

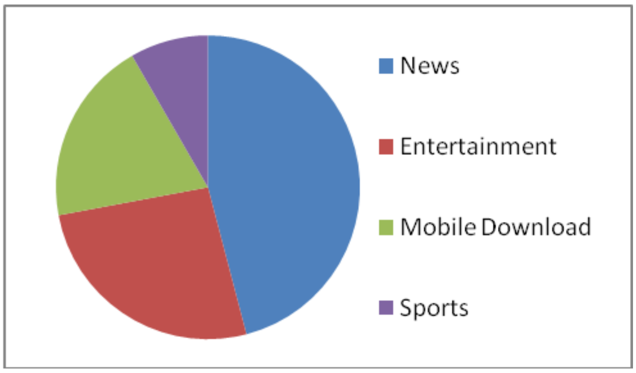

Figure 4.6(c). Proportions of mobile content recommendation from Dataset B.

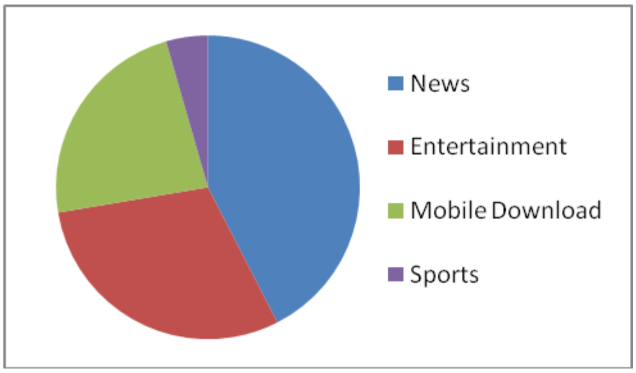

Figure 4.6(d). Proportions of mobile content recommendation from Dataset C.

4.4 Discussion on the Comparison Study

After establishing the integrated model for a mobile content recommendation system and the proposed method, MTCAR, a thorough comparison study was performed with other techniques to recommend appropriate mobile content that matches the needs of the user. In this chapter, the experiments have shown that MTCAR can perform well compared to other methods specifically finding relevant items after top six or top seven items, rules and recommendation generation.

Firstly, the standard measurements like accuracy rate and Mean Absolute Error (MAE) showed that MTCAR can perform better in terms of finding relevant items after the top six or top seven items, and it can also provide a similar accuracy rate for the first three items.

The reason why MTCAR provides a better recommendation is due to the MTCAR mechanism. It assembles clustering processes to identify user groups and predicts most wanted items from the cluster. Then, the relevant items are derived by association rules, which are generated from user clusters and target items in each cluster. So, relevant items would be created differently according to users' demographic factors and their different target items, whereas collaborative filtering (CF) is concerned with user ratings and focuses on finding relevant items or recommendations based on those ratings only. The rating is used to find the similarity of item. The same can be observed for association rules. This technique is helpful in terms of finding relevant items but the rules are constructed from user profiles only, which is not enough to consider item-based aspects. Therefore, it helps for the first three or four items, but in the later stages its performance on finding relevant items decreases.

Secondly, comparisons between MTCAR and association rules in terms of rules and recommendation generation showed that MTCAR returns highly acceptable results with the same support and confidence level. The number of association rules that are generated from MTCAR are less than the traditional association rules. That means MTCAR implements fewer rules to create a recommendation and gains better results. In addition, the number of items that can be recommended for the top 10 items is almost 10 items, while Association Rules has a limitation of 8 items on this measurement. Therefore, association rules is unable to recommend more items compared to MTCAR. Likewise, the level of emptiness, which means the recommendation system is unable to generate or recommend items to a user, also showed that MTCAR provides significant results. MTCAR gains much less in the emptiness level of a recommendation system compared to association rules.

Thirdly, when MTCAR is used with real world data, actual mobile content page view, the top mobile content categories derived from MTCAR in all datasets are similar to the mobile content page view with exception of a slight deviation from the page view in a couple of the items in Dataset C.

As stated above, it can be seen that MTCAR can be used in a mobile content recommendation system and it will provide better results compared to other techniques. This can address the limitation of the recommendation system for the first time user by recommending appropriate content that matches the user's needs. It also addresses first content rating in terms of finding relevant items. The proposed method, MTCAR, can enhance the mobile content recommendation system by its performance.

CHAPTER 5

MOBILE CONTENT RECOMMENDATION SYSTEM USING CLIENT-SIDE USER PROFILE FOR A REVISITING USER

5.1 Introduction

Having considered the non-interactive recommendation system in Chapters 3 and 4 for establishing the model for a first rater and a first time connection user, this chapter looks at the use of a client-side user profile for an interactive recommendation system. The assumption made in this chapter is that mobile users are willing to share some information when using the mobile devices so that the content can be customized to them. However, there are some users who may not like to rate the content items or participate in the profile collecting exercises. This group of users is not within the scope of our investigation. In this chapter, the users who are willing to share more information in the interactive recommendation system as revisiting users are classified. These kinds of users are proactive users and will interact with the recommendation system to make sure that the mobile content recommendation system can deliver contents that are close to what they are interested in. In order to deliver desired contents, a user profile with respect to the user's interest is to be stored and maintained. In this chapter, a client-side user profile that can be managed by the user is proposed.

5.2 Background of Using User Profiles

There are some works that demonstrate the storing of user's information on the mobile devices. MyLifeBits [26] stored everything related to a user's life in the user's computer. The research also suggested that mobile devices should also store personal data and the Internet usage behaviour of a user. In addition, the data should be organised according to the lifestyle and the time of the activities recorded. Digital content stored in a mobile phone can have emotional value [39], which can make a user more attached to the mobile device. In addition, it also found that historical usage data could be used for personalisation. In the Mobile Recommender System [124], the researcher has implemented a user profile management which can be run at the client's side. However, the Mobile Recommender System needs to install a shop application in order to manage a shopper's profile. The Mobile Recommender System was implemented as a non-interactive recommendation system. It cannot handle revisiting users and does not recommend relevant content items. In addition, the rating of products, movies and other content, are separated and the Mobile Recommender System is unable to deal with common interests or preferences.

By now it can be seen that use of user profiles to enhance the mobile recommendation system is needed. However, the structure of the user profile is an important area to handle. Due to the space limitation of a mobile device, appropriate structure should be considered. Only useful and important information should be stored for use by the mobile content recommendation system. The hierarchical structure of a user profile was proposed by Yun and Boqin [128]. They implemented a tag-based technique for a user profile and analysed the tags from website bookmarks. This can help to describe a target user's interests better. In addition, it shows that on average there are 2.65 tags per bookmark. A user profile for mobile content recommendation should include necessary information for the recommendation system to be able to identify which group of interests best matches the user's needs [21]. Age, gender, ethnicity and socioeconomic status showed that there are differences in the behaviour of using wireless devices [27]. A user's demographic factors should be collected in the user profile. Both the user's interests and the user's preferences are stored explicitly and implicitly in the user profile. They can be analysed to show an individual's usage pattern. In cases where a user's interests can be seen from the user's perspective, they are known as extrinsic preferences. A user can give information to a mobile content recommendation system about the content that they are interested in. The user can also rate those content items using some kind of information ranking. An example of incorporating a user's interests for a personalised search of information can be seen in Teevan et al. [129]. Their study focused on text-based personalisation for web searching. Germanakos et al. [130] mentioned that the building of the user profile could be complicated if the profile is to be incorporated with perceptual preferences. They also showed that interests and preferences are important factors to be used in a content recommendation system. It is assumed that the more information a user is willing to share, the more accurately the system can recommend the desired contents.

Moreover, user profiles should be allowed to update often in order to gain up-to-date recommended content. User profiles should not remain stagnant, but should adapt over time. For example, users may change their occupation and income which will affect how the user profile can be used. Hofgesang [131] mentioned that a user profile would collect information of the user as this information arises incrementally.

Hence, a user profile should be maintained and up-to-date. Jeon et al. [29] also showed an automatic updatable user profile used a user's preferences and interests for filtering mobile search. The user profile should be updated simultaneously every time users use the system. However, most research did not address the issues in the user profile of a non-interactive and an interactive recommendation system working concurrently. Hence, an integrated approach is needed, which is to establish an integrated mobile content recommendation system for first time user, first rating content and a revisiting user.

5.3 Flow Management for User Management on a Mobile Content Recommendation System

Before looking at the user profile for such an integrated mobile content recommendation system, the flow in the system needs to be examined. Figure 5.1 shows the steps for the first time user, where a user profile has not stored any information, progressing through the recommendation system. The user needs to add demographic information so that MTCAR can generate a recommendation. The user will then obtain a list of content items which have been recommended by the system. The user can choose to utilise content from this list or choose not to. After that, MTCAR might provide some feedback to the system and alter the user profile based on the content selected by the user. At this stage, a user can choose to be hidden and not interact with the recommendation system. If the user chooses to be anonymous, that user will not receive personalised content.

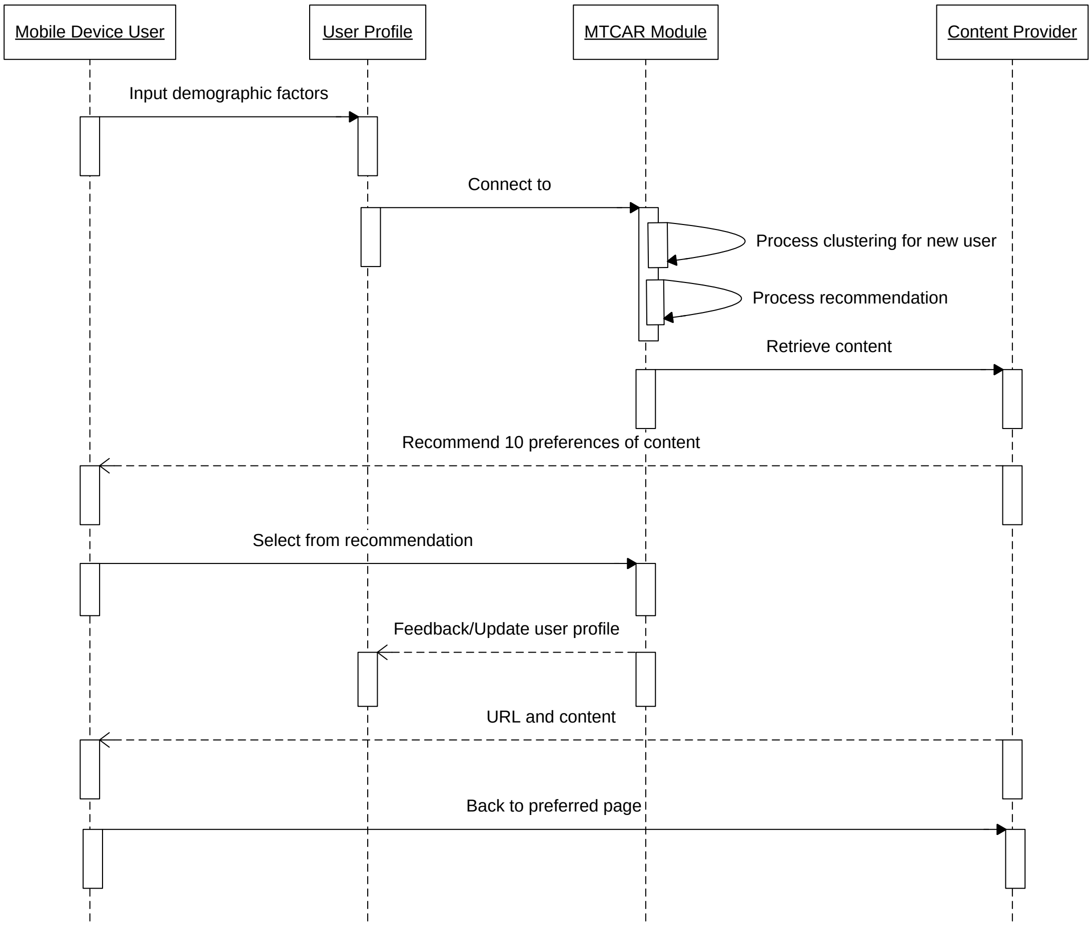

Figure 5.1. Sequence diagram of steps for the first time connection user with the process of mobile client-side user profile and content provider.

Having considering the first time user, the next group of users using the mobile content recommendation are the revisiting users. A revisiting user refers to the user in the recommendation system that has experienced the system and their updated user profile due to increased participation with the system. Unlike a first time user, there is stored information, including extrinsic and intrinsic interests, and preferences. When the user connects to the mobile Internet, the updatable user profile will upload the updated information, using MTCAR for more personalised recommendations. Next, the type of content selected and utilised by the user will also be stored in the user profile under the Common Content. This can store the information on how the user utilises the content in the mobile content recommendation system. Figure 5.2 presents the sequence diagram of a revisiting user.

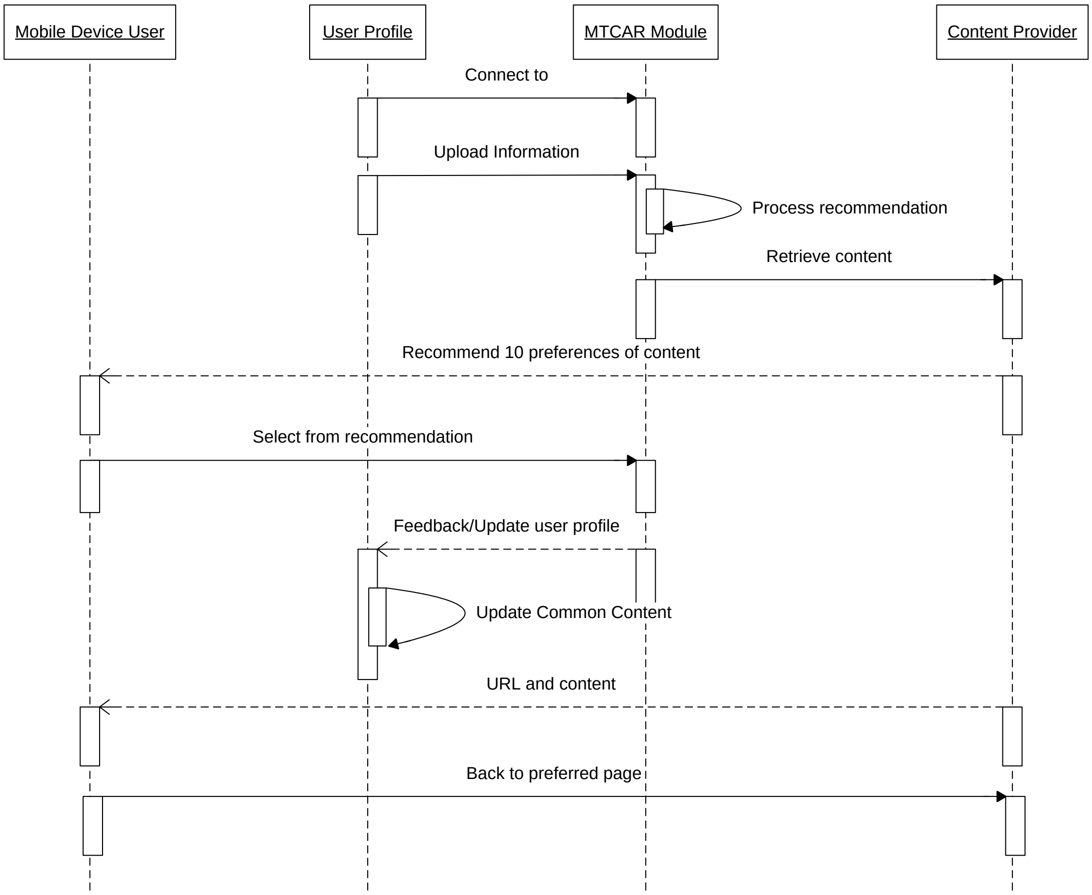

Figure 5.2. Sequence diagram of steps for a revisiting user for a mobile content recommendation.

The benefit to a revisiting user is that the user can obtain appropriate content that matches one's interests and preferences in a user profile. This also includes relevant content items from the recommendation system.

5.4 Client-side User Profile Prototype

5.4.1 Basic Functions and Personal Information Management

This section presents the mobile client-side user profile prototype and basic functions for the user profile. Figure 5.3 shows the main menu of the user profile where the user can manage his or her information, including preferences by selecting from the menu.

Figure 5.3. Client-side user profile main menu.

In Figure 5.4, it can be seen that users are able to input their basic information of the demographic factors and this information will be stored at the client's side. The user has a right to change or alter the information, with regards to his or her level of participation with the recommendation system.

Figure 5.4. User's client-side mobile personal information and demographic factors.

5.4.2 Privacy of a Mobile Content Recommendation System

Privacy is also an important issue that users have concerns about when they go online, especially on the mobile Internet. Users may be concerned about the possibility of their information being stolen. Therefore, it is a fundamental issue that must be addressed. Privacy can be considered an issue that governs the flow of information, which is outgoing and incoming. It should be managed and decided by the user as follows:

- **Outgoing allowing:** The system will ask a user for consent for personal information to be submitted to a destination server. The user has a right to determine whether that the information can be sent or not sent.
- **Incoming allowing:** It is similar to outgoing allowing. The user will be informed that the user profile on a mobile device will be modified. The user will also be informed that some information from the server will be updated in the user's mobile device.

This can help a user manage privacy issues by allowing the user to make decisions. It should also be easy for the user to understand the process as well. This concept of the privacy issue is applied from the user-centric privacy framework [132]. This can increase the user’s confidence with regards to the privacy issue due to the way it is handled by the recommendation system.

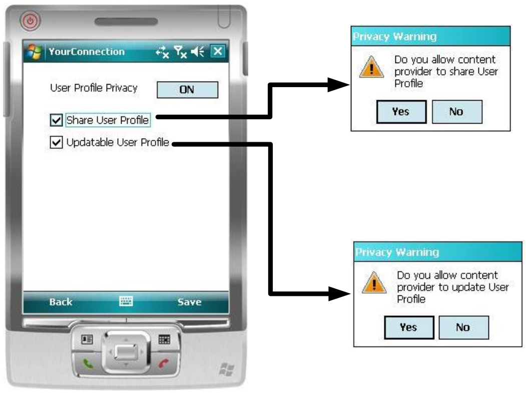

Figure 5.5. User profile privacy management.

The prototype of a mobile client-side user profile also addresses user privacy issues. The user profile privacy should be a feature that can be turned on or off. This feature will determine whether user profile information is shared or not shared with the content provider. In addition, the user has a right to select options for sharing the user profile with the content provider and to allow the content provider to update the user profile as well. Warning message dialogues will also be displayed for each option selected, to ask the user whether the user wants to share or update user profile information with regards to privacy. Therefore, through the user profile management the user has a right to consent to sharing information with the content provider and allowing the content provider to update user profile information.

5.4.3 Client-side User Profile Structure for a Mobile Content Recommendation

The client-side user profile not only includes user demographic factors but it also stores content preferences which are separated by content providers. Within the stored content provider preferences, content and its ratings are stored. The user can view and edit values for content and ratings for each content provider. The client-side user profile counts the number and types of content being accessed. The client-side user profile then assigns a rating for that content. The client-side user profile assigns a higher rating to content which is accessed multiple times. For example, if the user frequently accesses economic news content, economic news content will be assigned a higher rating. For content that the user has not rated but has accessed, the client-side user profile is updated and this content will be counted and the rating for this content will increase automatically. In addition, although content provided by each mobile content provider might be different, some content might share common features. Common Content, stores content and ratings are based on how many times specific content is accessed. For example, if a user accesses content with investment information from three different content providers, "investment" will be recorded to Common Content. Therefore, Common Content in the client-side user profile will deal with common content information as an intrinsic preference. The user can manage Common Content features as well.

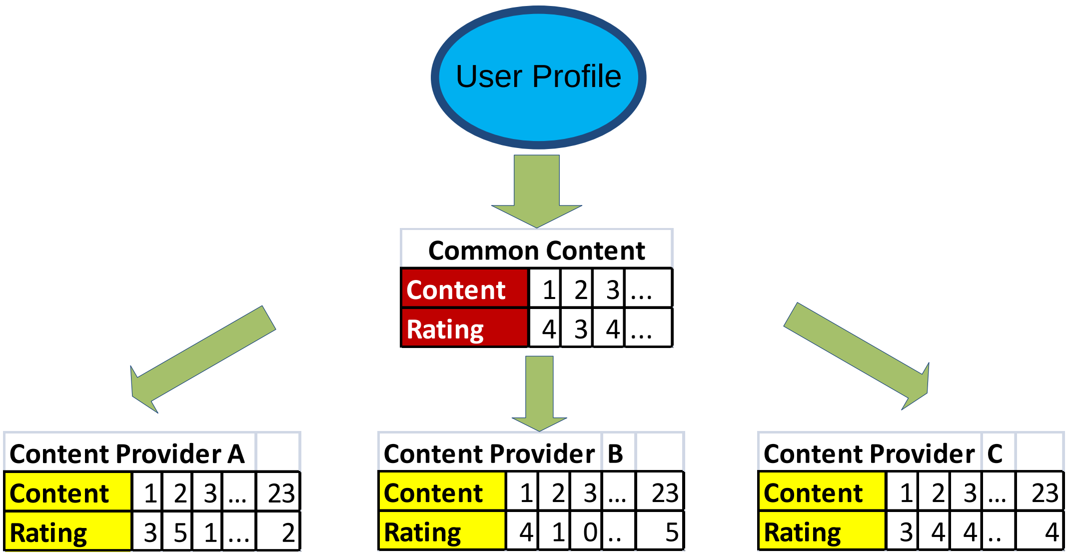

Figure 5.6. The structure of a mobile client-side user profile and Common Content.

To communicate with the content provider, a client-side user profile is also implemented using standard programming language, which is XML, in the user profile structure. XML encourages interoperability features on the user profile. This can ease communication between the user and the content provider. This also helps the content provider to process the recommendation lists more efficiently. Figures 5.7 and 5.8 show examples of XML where the client-side user profile communicates with the content provider through Common Content. The client-side user profile also communicates content rating for each content provider's content.


```
<?xml version="1.0" encoding="UTF-8" ?>
- <CommonContentsRating>
    <CommonContents.content_id>7</CommonContents.content_id>
    <content_name>car</content_name>
  - <keywords>
      <keyword id="1">car</keyword>
      <keyword id="2">automobile</keyword>
    </keywords>
    <common_rating>5</common_rating>
  </CommonContentsRating>
```


Figure 5.7. Present example of Common Content in XML format to communicate with content provider.


```
<?xml version="1.0" encoding="UTF-8" ?>
- <websiterating>
    <URL>http://bmw.mobi/bmwmobi/sx38ef27c6xs/</URL>
    <cluster>5</cluster>
    <website.website_id>1</website.website_id>
    <contents.website_id>1</contents.website_id>
  - <contents>
      <content_id>7</content_id>
      <rating>4</rating>
      <content_id>13</content_id>
      <rating>5</rating>
      <content_id>33</content_id>
      <rating>4</rating>
    </contents>
  </websiterating>
```


Figure 5.8. Present example of website rating stored at client-side in user profile in XML format to communicate with content provider.

5.4.4 Client-side User Profile Preference and Revisiting Management

The user has a chance to manage his, or her, own preferences in order to obtain personalised content and get recommendations, which are related to their preferences. Figure 5.9 shows user preferences can be adjusted, to rate content based on his or her preferences. In addition, in this prototype, the user can add types of content and related keywords, which can be updated automatically to client-side user preferences when the user contacts the content provider and receives feedback. Selecting the menu button reveals an edit function to the user for editing content types, content keywords and content ratings, according to the user's needs. This can be seen from Figure 5.10.

Figure 5.9. User's preference for client-side user profile management.

In addition to the user profile, using XML appears more user-friendly than the traditional profile or storing information using 'cookies' in terms of ease of understanding because the user profile is readable and stored at the client-side. Users can understand this method of storing information and they can manage this information by themselves.

Figure 5.10. Adding and editing of preferred content to a user profile preferences management system for the client-side user profile.

Revisiting users are the users who re-access content from a content provider. They might have some additional information to share, such as preferences from mobile Internet surfing. Therefore, having a feature to manage Common Content for a revisiting user is also important because it allows the accumulation of user data and information as the user keeps accessing mobile content. This helps the mobile content provider to determine the most appropriate content for the user and deliver the content to that user. Therefore, the feature of managing content from a content provider should be available as a basic preference similar to managing preferences in Common Content. This might lead to more appropriate recommendations for the revisiting user.

The revisiting user can view how many times he or she has accessed the content, as well as the ratings for both Common Content and each content provider. Users can manually manage their user profile, in order to obtain appropriate recommendations for mobile content. In addition, users are able to manage preferred content for each content provider that they revisit, so that each content provider can provide suitable content according to the user’s needs. The type of the content is also recorded to Common Content and client-side user profile preferences. This means if content appears in a content provider’s list, it will also appear in Common Content. The example of revisiting website content management can be seen from Figure 5.11

Figure 5.11. Example of revisiting website content management.

The client-side user profile will attempt to implement content-based filtering (CBF) and its management for the mobile content recommendation system. This depends on how users share their information. The content provider can then utilise this information to recommend content to the users. In addition, this shared information is also useful for other users when performing collaborative filtering for content. Figure 5.12 shows the overview of the user profile management at the client’s side.

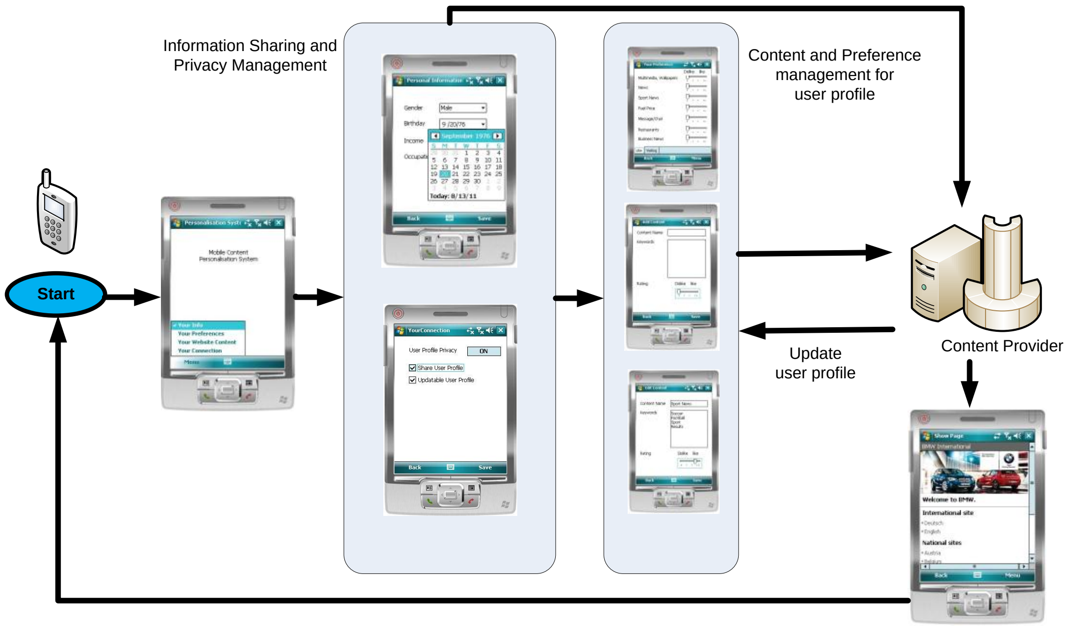

Figure 5.12. Overview of client-side user profile management.

5.5 Discussion and Conclusion

The user profile is the important factor in a recommendation system to help a user obtain personalised content and obtain recommended relevant content items. It is a trend for mobile device to store user's information within the device and encourage mobile content filtering. However, there is a technological change in the mobile device rapidly. Therefore, further work related to structure and schema of user profile is required. In addition, the user profile also assists the content provider to deliver appropriate content that is suitable for the user's needs at that time. If the user shares some information, it enables the recommendation system to recommend appropriate content. However, users may be concerned about what information is shared; therefore, profile management at the client's side seems to be an appropriate solution for this scenario.

In this chapter, the prototype of client-side user profile management is demonstrated. It also demonstrates how personal information is stored and content is managed at the client's side. In addition, the user profile can be updated automatically by the content provider and manually edited by the user. Privacy issues are addressed in this prototype as well. This is achieved by incorporating features which handle privacy issues through user consent and permission. The recommendation system can provide better recommendations to revisiting users when these users share more information in the user profile.

CHAPTER 6

USING TIME CONTEXT INFORMATION FOR MOBILE CONTENT RECOMMENDATION

6.1 Introduction

The characteristic of a mobile user is the mobility features during various times of the day. Moreover, the user needs the content provider to respond to them immediately when they connect to the server in order to access the right content at the right time. In this chapter, the use of time context information is incorporated into the mobile content recommendation system. To narrow the discussion and focus of this thesis, only an adaptive list menu with respect to the time-of-day will be investigated.

The study of Partridge and Golle shows that time is the one of the important factors for context awareness [133]. This research explores a factor of time which is related to personalisation and an adaptive menu of content recommendation for a mobile Internet user. The menu can be varied and the option list rearranged according to a user’s need based on the time-of-day. The most desirable items of the list menu during that period when a user connects to the mobile Internet will be shown. This can reduce several clicks on the mobile when searching for the desired content at a specific time of the day.

In this chapter, the use of time for a recommendation system will be studied, and the case study is based on a group of users who download mobile games. It is assumed that this group of mobile Internet users has been clustered using methods described in Chapters 3 and 4. The focused domain in this case study is mobile games.

In recent years, it can be observed that there is a rising trend of mobile games developed for a huge range of mobile devices. During different times of the day, mobile device can be used as an entertainment device for relaxing or while waiting for buses or before an appointment. A computer game has been viewed as an interactive media that was played most often when people had free time. Therefore, it is easy to realize the impact by combining the mobile phone culture and entertainment games. This leads to the popular trend in developing mobile games. Many mobile games have been developed in recent years. Some even transform popular console or PC games from the past to a mobile platform. However, there could be too many games for mobile users to select. Additionally, the users may prefer different types of mobile games at different times of the day. When users download a game, they may spend too much time browsing for content. Furthermore, many users have to pay for the game. Although some mobile games provide demo or trial versions, the users may also spend too much time browsing for something they like.

In this proposed recommendation system, the downloaded games should be considered, as well as a user's expected playing period. The theme of the game should also be taken into consideration. Eventually, the recommendation system will configure the game downloading menu via the mobile Internet. In addition, the time and the theme of a game can be used primarily to predict a user's needs. This will greatly reduce the browsing time the user spends searching for the type of games he/she prefers. In order to solve this kind of problem, filtering and recommendation concepts can be applied to generate the menu for downloading mobile games.

In this chapter, a mobile game recommendation system by using time-of-day and day-of-week information, and the theme of the game, to cluster and rearrange the most appropriate menu is presented. This work also tries to reduce a user browsing several levels of menu in order for the user to find what they want. Whenever a user downloads the game, the menu which is able to meet the user's needs according to the time period and the day-of-week, as well as the game theme, will be displayed on the top of the menu. A user can easily click to download or play the most desired game. The system also identifies the period used during a weekend and the time-of-day during that time.

6.2 Research Methodology

This section describes the experimental design and a process of experiment.

6.2.1 Experimental Design

This research used the server log file of a mobile content provider in Thailand which provided several type of content related to entertainment for a mobile phone, including Java Games, Theme, Wallpaper, Ring tone, Video clip, et cetera. The log file gathered 9,644 unique users and 60,000 transactions. In addition, the log file also recorded the content name and content category of the company's file server and there were date and time usage information as well.

When a user connects through a content server, the content visitor page will appear. Next, the content visitor page, category visitor page, list visitor page and detail visitor page will be accessed respectively. There are two types of customer, member and non-member. A member customer can download unlimited content by monthly payment while a non-member has to pay a per-time charge when downloading the content. Additionally, the member can click to download the content via a list visitor page, but a non-member should download the content by passing through a payment and charge page. Both types of customer session usage or any pages access will be recorded in the server log file.

The member type customers can download unlimited content including Java games by monthly subscription. Therefore, this type of member is not too concerned about the number of downloaded games. However, they may be concerned with the difficulty in finding their game of interest in the shortest time, due to the data charges. In contrast, non-members would be concerned about both the connection fee and finding their game of interest fast. In Thailand, there are more pre-paid customers than post-paid customers. The pre-paid customers seem to be more restricted with their budget than the post-paid customers as well.

Half of the data was divided for training and adding time-separation factors with 4,467 unique users, including 29,998 transactions. The other parts of the data would be used to test the result. However, in this log file, the user name or any user profile cannot be collected and kept at the server. The only data that can identify a user is the mobile phone number and this cannot be disclosed.

The design of this experiment started with pre-processing data stage. From the server log file, the data cleaning method was applied to obtain appropriate data format for analysis. Then, it was imported to a database for the experiment. After that, the factor of the time-of-day was classified in order to know which item would be used in exactly which period. The periods were separated in the following parts specifically: 1) 5:01-11:00 was assigned as Morning; 2) 11:01-16:00 was assigned as Afternoon; 3) 16:01-21:00 was assigned as Evening; and 4) 21:01-5:00 was assigned as Midnight. Next, the period was appended and classified into the data according to the time-of-day, the classification method being by sorting out the seven items most frequently used during that period.

After that, the irrelevant data should be removed, as some models cannot download the content so the content name and content type in the database would be left as a blank data. When the data was filtered again, the remaining data can be used to analyse by categorisation corresponding to the period. It can be divided into four periods and kept as follows:

$$P_i = \{C_1, C_2, C_3, \ldots, C_7\}$$

where P_i is the set of top seven content names at the period *i* and C_n is the content name of the mobile content page which was ranked according to the most frequently used at the period *i*.

Then, the test data was organised in the same way as the training data by removing irrelevant records which cannot be downloaded by that mobile device. Next, the classification of a mobile Internet usage session was managed by grouping the users who were using the mobile Internet at that time. It was separated user by user and session by session as follows:

$$US_j = \{C_1, C_2, C_3, \ldots, C_n\}$$

where US_j is the user who connects through the mobile Internet each session and C_n is the content page identification or content name which was used in session *j*.

In addition, the user session was referred to a record of user clicks and content page usage at the time a user connects to the mobile Internet for one period. The user may access one content page in the session or several pages. It depends on how easy the user can find the most wanted content or not. For example, the user may access only one page if the page and content were displayed at the first mobile webpage. In contrast, it may take a long time as the user may need to drill down several levels if the user cannot find the desired content in the pages that appear.

The accuracy rate was calculated by counting the number of matching content names between a period and a user's session with the testing data. If there is at least one content name in the period content set (P_i) which is directly matched to the content in US_j, it was counted as 1. The overall sessions would be calculated whether there was at least one content matching or not. It can be explained as follows:

$$\textit{Number of matching session} = \sum_{j=1}^{n} \begin{cases} 1 : P_i \cap US_j \neq \varnothing \\ 0 \end{cases}$$

6.2.2 Experimental Process

The pre-processing phase is first carried out by gathering the log file. Then, the data was transferred to a database for the convenience of issuing queries. The data cleaning process is also an important process in this phase. There could be much irrelevant information in the log file and this process will eliminate that unwanted information. For example, users who were unable to download the game were removed from the database. Data conversion and formatting were performed as well. In the main process, classification of data based on the context information was implemented. Each important factor was used to classify the data in each group. The results were managed and sent to the user as a personalised downloadable game menu, which will correspond to the user's needs based on the time information. This can be seen from Figure 6.1.

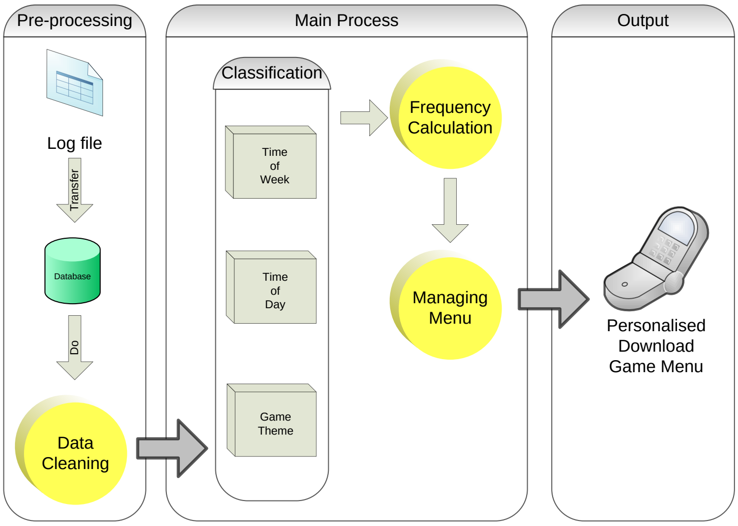

Figure 6.1. The downloading game menu process overview.

6.3 Experimental Results

The results that are shown in this section are user click for time-of-day periods, game downloading with classification factors and proposed downloading mobile game menu.

6.3.1 Results for User Click for Time-of-day Periods

The main results are shown for the experiment. Table 6.1(a) to (d) shows the result of classified content name, content type and content category with an accumulated frequency of a user's click. As the data is confidential and cannot disclose the name of the content, the content name will be altered to just content ID instead.

Table 6.1(a)

The Contents are Ranked by the Frequency of the User's Click According to Period 1 (P_1)

Content ID	Type	Category	Frequency
Game 33	Java Games	Action	106
Ring tone 7	True Tone	Sport	92
Game 133	Java Games	Card&Casino	31
Game 129	Java Games	Action	25
Game 204	Java Games	Action	25
Game 250	Java Games	Action	20
Game 251	Java Games	Action	17

Table 6.1(b)

The Contents are Ranked by the Frequency of the User's Click According to Period 2 (P_2)

Content ID	Type	Category	Frequency
Ring tone 7	True Tone	Sport	401
Game 33	Java Games	Action	322
Ring tone 14	True Tone	True Tone	117
Game 133	Java Games	Card&Casino	85
Game 250	Java Games	Action	70
Game 251	Java Games	Action	68
Theme 24	Theme	Cartoon	57

Table 6.1(c)

The Contents are Ranked by the Frequency of the User's Click According to Period 3 (P_3)

Content ID	Type	Category	Frequency
Ring tone 7	True Tone	Sport	1527
Game 33	Java Games	Action	755
Ring tone 14	True Tone	True Tone	307
Game 133	Java Games	Card&Casino	243
Theme 24	Theme	Cartoon	209
Ringtone 13	True Tone	True Tone	184
Game 250	Java Games	Action	172

Table 6.1(d)

The Contents are Ranked by the Frequency of the User's Click According to Period 4 (P_4)

Content ID	Type	Category	Frequency
Ring tone 7	True Tone	Sport	1630
Game 33	Java Games	Action	729
Ring tone 14	True Tone	True Tone	326
Game 133	Java Games	Card&Casino	241
Theme 24	Theme	Cartoon	227
Game 204	Java Games	Action	203
Game 250	Java Games	Action	186

The next result is related to the accuracy rate of the personalised adaptive menu according to the time-of-day of the content page compared with the user's session. The list-oriented menu will be changed by bringing the top seven from the P_i to display on the first page. If there is at least one content that matches users' usage in their session, the accuracy rate will be increased. As can be seen from Table 6.2, the percentage of matching sessions was higher than 77% and it reached 81% in Period 2 which is the most number of users' sessions compared with other periods. In Overall 1, its number of users' sessions was higher than the number in Overall 2, because some users might use the mobile Internet in overlapping time-of-day periods. For example, user ID 3233 might use the mobile Internet from Period 2 to Period 3. As a result, the accumulative number in the overall came from the sum of the sessions in all periods, while Overall 2 presented the distinct user ID for all periods and compared it with the unique session as well. The accuracy rate for the distinct user ID could reach the accuracy rate of 80.81%.

Table 6.2

The Accuracy Rate of Proposed Personalised Adaptive Menu

Period	1	2	3	4	Overall 1	Overall 2
user's session	730	858	589	1044	3221	2902
match session	573	695	469	804	2541	2345
Accuracy rate (%)	78.49	81.00	79.63	77.01	78.89	80.81

Moreover, the results also presented the ranked content in each period. In Period 1 (Morning), the users aim to download Java Games more than other content type; six out of seven of the top ranked games were Java Games in this period, while in other periods the top downloaded content types were varied, such as Java Games, True Tone or Theme. It can imply that users have little time in the morning to alter their mobile phone profile. They need only relaxing content like games while the users in other periods may need to alter the mobile phone profile by downloading content relating to the device, such as ring tones or themes.

6.3.2 Game Downloading with Classification Factors

Three main results related to the factors are tabulated in the tables. Table 6.3 presents that the action game is the most favourite game theme with the time-of-week factor. It reached around more than 70% compared with other themes. The following favourite mobile game themes are casual, puzzle and adventure games at 10.99%, 7.03% and 5.30% respectively. The rank seems to be the same when compared across the weekday, except the percentage of puzzle and adventure games which have a higher percentage. It can be seen that the RPG game does not generate much interest as a game theme for mobile games due to the longer storytelling and the time needed to complete the game. Therefore, the downloading game menu during weekdays may add more games on the puzzle theme.

Table 6.3
Time-of-week and Game Theme Factors

Game Theme	**Weekend**		**Weekday**	
	Frequency	Percentage	Frequency	Percentage
Action	5162	72.53%	7789	70.52%
Adventure	377	5.30%	619	5.60%
Casual	782	10.99%	1181	10.69%
Others	9	0.13%	36	0.33%
Puzzle	500	7.03%	872	7.89%
RPG	14	0.20%	16	0.14%
Strategy	95	1.33%	180	1.63%
Sports	178	2.50%	352	3.19%
Total	7117	100.00%	11045	100.00%

The next result used the time-of-day factor which was also separated by time-of-week to rank the most downloaded game themes. It can be seen from Table 6.4 that there are different proportions of downloaded game themes in each period of time-of-day and in the time-of-week as well. The results were shown in the table below.

Table 6.4
Time-of-day and Time-of-week Factors

Time	**Weekend**								**Weekday**							
	1		**2**		**3**		**4**		**1**		**2**		**3**		**4**	
Game Theme	**Freq.**	**%**	**Freq.**	**%**	**Freq.**	**%**	**Freq.**	**%**	**Freq.**	**%**	**Freq.**	**%**	**Freq.**	**%**	**Freq.**	**%**
Action	779	75.63	1532	75.17	1344	68.61	1507	72.11	1281	72.50	1862	71.64	2435	71.81	2211	67.24
Adventure	66	6.41	84	4.12	131	6.69	96	4.59	100	5.66	156	6.00	190	5.60	173	5.26
Casual	90	8.74	220	10.79	244	12.46	228	10.91	150	8.49	268	10.31	339	10.00	424	12.90
Others	4	0.39	1	0.05	1	0.05	3	0.14	6	0.34	6	0.23	8	0.24	16	0.49
Puzzle	51	4.95	133	6.53	151	7.71	165	7.89	118	6.68	196	7.54	261	7.70	297	9.03
RPG	4	0.39	6	0.29	3	0.15	1	0.05	3	0.17	0	0.00	8	0.24	5	0.15
Strategy	27	2.62	42	2.06	55	2.81	54	2.58	79	4.47	80	3.08	89	2.62	104	3.16
Sports	9	0.87	20	0.98	30	1.53	36	1.72	30	1.70	31	1.19	61	1.80	58	1.76
Total	1030	100.00	2038	100.00	1959	100.00	2090	100.00	1767	100.00	2599	100.00	3391	100.00	3288	100.00

After the time-of-week, time-of-day and game theme factors are calculated and sorted, the result of the proportion of games downloaded via the mobile phone is obtained. The game themes which attract more interest are action, casual, puzzle and adventure games. Importantly, if the menu used the result from action games only, the other game themes would have less opportunity to be downloaded by the customers. Thus, the proportion of the favourite game will be used as a share of the menu items on the mobile game menu according to the time-of-week and time-of-day. Then, each game theme contains its game rank, and recommendation of the more favoured games will be displayed on the main downloading game menu. There is a quota for the providers to offer their games in each game theme to the customers. For example, action games will be given the top seven games of its recommendation list while casual, puzzle and adventure games have one game for their quota. In other word, they can send their top ranked game to be the candidate in the main downloading menu to increase profit.

6.3.3 Proposed Downloading Mobile Game Menu

This section presents proposed downloading mobile game menu according to time-of-week and time-of-day combined with time-of-week.

6.3.3.1 Proposed Menu According to Time-of-week

From the result using the time-of-week to cluster, it appears indifferent when it uses only the time-of-week to rearrange the menu. However, this result also increases more chance of the downloaded appearing game by giving the quota to bring the top rank in each category to the main downloading menu. Otherwise, if frequently downloaded games only are used, the games in other game genres would not be downloaded at all. Furthermore, this proposed menu can be used and implemented for the personalised downloading mobile game system. It also facilitates a user to downloading the games they are more interested in, according to their needs. Due to confidentiality, the source of information would not like to disclose the actual game's name. Therefore, the game ID will be used in this research instead of the game's name. The proposed menu using time-of-week factors can be seen from Table 6.5.

Table 6.5
Proposed Menu Using Time-of-week Factors

Weekend/Weekday	
Game-id	**Game Type**
Game-041	Action
Game-160	Casual
Game-310	Action
Game-253	Action
Game-155	Action
Game-311	Action
Game-252	Action
Game-067	Adventure
Game-048	Puzzle
Game-245	Action

6.3.3.2 Proposed Menu According to Time-of-day Combined with Time-of-week

As can be seen from Figure 6.2, the graph shows that the action games during Period 3 of the weekend tend to decrease but the percentages of casual games increased, as well as puzzle games. The percentage of the adventure games fluctuated throughout the periods. As a result, the weekend downloading game menu for the third period can be altered from the normal menu using time-of-week factors. When the quota of action games is decreased, the competition among other game themes will occur. The most frequently downloaded games at that time will have the place to show the downloading item or its name on the game menu. For example, in the experiment, action games gain six places out of ten while the others gain one place. Therefore, the second ranked games in each remaining category are compared to find the maximum frequency in order to gain that place. The result shows that the second ranked game which has the most frequency is in the puzzle game category. It can be seen from Table 6.6 that in Period 3 of the weekend at least five out of ten games ranked on the downloading game menu are different from the proposed menu using time-of-week factors. Compared to Period 4 on the weekend, the menu order is different from Period 3 and a little bit different from only the time-of-week. Nevertheless, this can facilitate the user to download the game they are more interested in to their mobile phone according to the time-of-week and time-of-day. It can reduce the click distance to find the desired game and reduce the clicks needed to scroll down several levels of content page to find the desired games.

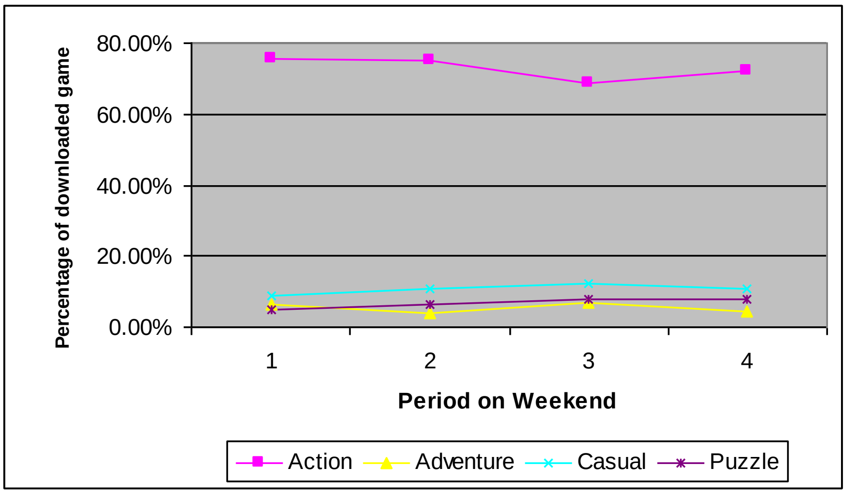

Figure 6.2. The trend to download games over the time period of the weekend.

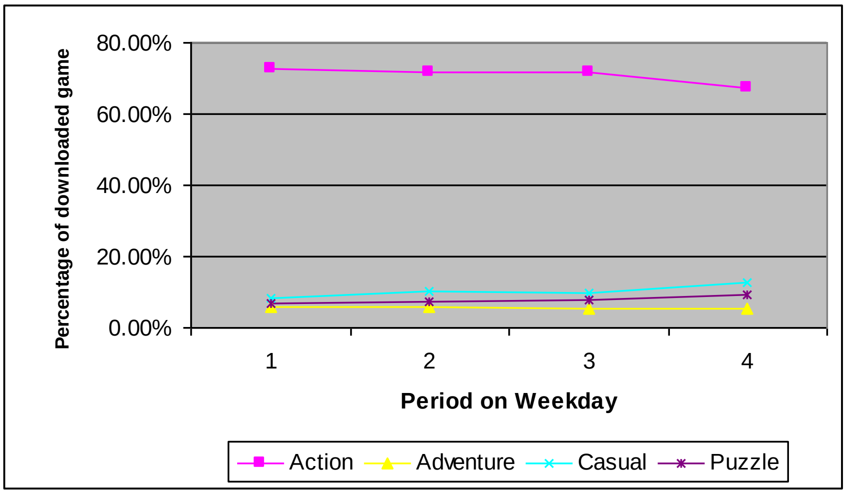

Figure 6.3. The trend to download games over the time period of the weekday.

Table 6.6

Proposed Menu Using a Combination of Factors on the Weekend

Weekend Period 3 Menu		Weekend Period 4 Menu	
Game ID	**Game Type**	**Game ID**	**Game Type**
Game-041	Action	Game-041	Action
Game-160	Casual	Game-160	Casual
Game-310	Action	Game-310	Action
Game-253	Action	Game-253	Action
Game-067	Adventure	Game-155	Action
Game-155	Action	Game-311	Action
Game-311	Action	Game-252	Action
Game-252	Action	Game-048	Puzzle
Game-048	Puzzle	Game-067	Adventure
Game-068	Puzzle	Game-245	Action

On the weekday, it can be seen from the Figure 6.3 that the downloading percentage of action games has declined in the fourth period markedly. In contrast, the percentage of casual games has increased, as well as the puzzle games. Table 6.7 shows the proposed menu using a combination of factors on weekdays.

Table 6.7

The Proposed Menu Using a Combination of Factors on Weekdays

Weekday Period 4 Menu	
Game ID	**Game Type**
Game-041	Action
Game-160	Casual
Game-310	Action
Game-253	Action
Game-067	Adventure
Game-155	Action
Game-311	Action
Game-252	Action
Game-048	Puzzle
Game-068	Puzzle

6.4 Discussion

This chapter proposed a method which facilitates mobile device users when they connect to the mobile Internet using time context information. The group of mobile Internet users in these experiments is focusing on the mobile game downloader.

Due to the increasing number of mobile games being downloaded via the mobile Internet, it is getting more important to provide a more personalised experience. It can be observed that there are several game genres preferred during different periods and times, depending on situations. In order to provide faster and more personalised service, rearranging the game downloading menu seems to be important for the users. The system makes use of the server log file of a mobile game provider in Thailand, specifically feature phone user, to prove the concept.

This chapter showed the results on mobile game downloading from introducing a mobile game recommendation system using time-of-week and time-of-day information. The system allows users to access the content they are interested in without scrolling down several levels of menus and pages. Furthermore, this can help a user reduce the connection time of the mobile Internet as well. As can be observed from the results, it shows that our assumption of the factors that should be included when creating the personalisation mobile game recommendation system is valid. If the contents were separated by period using the time-of-day, users' menus can be adapted to suit users' needs. The results show that an adaptive menu reached an accuracy of 80.81% and each period provided matching users' sessions at around 77%. Content category can be varied in each period of the time-of-day. It can be inferred that users will use different categories in different periods of the day. Moreover, the results also show that the time-of-week can provide a different adaptive menu to a user and present different game downloading options.

In conclusion, this chapter drills down in detail after a user group has been formed by clustering. After that, the predicted content has been classified including association rules are done. Then, the context information element of time is investigated for mobile content recommendation. It can be seen that for specific content, like mobile games, time can be the factor to determine the content that a user may need in different contexts. In the feature phone, users may use different times in their daily life to use a phone. Moreover, the browsing content is also different as well. As a result, the personalised downloading menu or recommended content should be altered according to the different time periods.

CHAPTER 7

CONCLUSIONS

7.1 Research Summary

This thesis investigated the problems that exist in mobile recommendation systems and developed an integrated approach to address the current limitations, especially in the early stage in terms of the first time user connection and first rating items. It includes the prediction of relevant items for a mobile content user to enhance the mobile content recommendation system. The established integrated model also makes use of the user's profile for any new user connected to a mobile recommendation system.

In addition, the integrated mobile recommendation system uses a client-side user profile where the user profile can be self-managed by new users and revisiting users. The recommendation system is also enhanced by using dynamic time context information. It can predict dynamic recommended content based on mobile context information.

An integrated mobile content recommendation system has been developed to address the problem of a new user connection or a first time user. Both types of user are referred to as the user with insufficient information to create mobile content recommendations. This integrated model handles this kind of user using a non-interactive recommendation system with multiple processes. This system also addresses the problems of first rater items for a non-interactive recommendation system. The proposed methodology has covered recommendation processes for mobile content filtering where processes are referred to prediction and recommendation. It showed the process that can handle a first time user connection. In addition, the method not only predicts top content items but also retrieved relevant items. This can enhance the performance of a mobile content recommendation system. The important aspects of the integrated mobile content recommendation system are summarised as follows:

1. The integrated model for a mobile content recommendation system starts with determining a mobile user group. Clustering analysis helped the recommendation system to classify a mobile content user group based on criteria and demographic factors. Thus, each cluster was identified to be a typical mobile user group for a new user or an active user in the recommendation system. However, there is a challenge working on clustering for a mobile content user with partition clustering. It is the problem of determining the optimal number of mobile user clusters. Therefore, the proposed Zoning-Centroid method helped to determine the number of clusters to facilitate cluster analysis for mobile content recommendation. The Zoning-Centroid method does this by performing a separation distance measure between the cluster centre and the data into the zone in order to form the cluster. This could help unsupervised learning data to identify the classes. The determining of the user group is to address the limitation of mobile content recommendation in the early stage. In addition, it can also cope with the problem of a first time user.

2. Mobile content filtering for the top content items incorporated classification techniques which could help mobile content recommendation to determine the predicted top content items. In this stage, different classifiers have been compared and evaluated by the proposed measurement, CMScore, for identifying the top content items. By using the CMScore, it can determine which classification method is best used for the cases under investigation. Decision tree with C5.0 algorithm is appropriate classification technique for the cases under investigation using CMScore.
3. Predicting the relevant content items for mobile content recommendation is also an important issue to be handled for first rating content. This can enhance the mobile content recommendation system such that the user will be given a chance to browse through new or un-rated items. The proposed MTCAR facilitated rule generation with level separation and target determination by top items in the mobile content filtering phase. MTCAR can also extract rules by reducing rule complexity and eliminating redundant rules by the rule consolidation process.
4. The proposed integrated model for a mobile content recommendation system, MTCAR, has also undergone some comparison studies with some popular methods using for mobile content recommendation system. MTCAR extracted fewer rules and assembled the clustering and classification mechanism to determine a group of users and top content items used in the group. Therefore, MTCAR can perform better compared to other methods in terms of finding relevant content items above top six or seven items. The results also showed that MTCAR could predict more relevant items than

other methods. In addition, the rules extracted by MTCAR are much lower compared to association rules. MTCAR is also verified when compared with real world data with top mobile content categories. MTCAR can recommend content items similar to a mobile content page view with the exception of a slight deviation from the page view in a couple of the items in Dataset C.

5. The user profile is also an important factor in a recommendation system to help a user obtain personalised content and have relevant content items recommended. In addition, the user profile also assists the content provider to deliver appropriate content that is suitable for the user's needs at that time. The recommendation system can also provide recommendations to revisiting users, when these users share more information in the user profile that is stored and managed at the client's side.
6. The dynamic time information was added to enhance the mobile content recommendation system. The case study presented in the thesis is based on mobile game downloading. It can be observed that there are several game genres preferred during different time periods. In order to provide faster and more personalised service, rearranging the game downloading menu is important for the users. The results demonstrated that mobile game downloading becomes more personalised by incorporated time-of-week and time-of-day information in the integrated mobile game recommendation system. If content was separated by time periods using the time-of-day, the mobile device menu can be adapted to suit user needs. It can be inferred that the user will use different categories at different times of the day. Moreover, the results also show that the time-of-week can provide a different adaptive

menu to the user and recommend different games for downloading at different times. Therefore, the personalised downloading menu or recommended content should be altered for different time periods.

7.2 Future Work and Directions

The integrated mobile recommendation system in this thesis has provided a framework towards a general purpose mobile content filtering for recommendation systems that can handle some of the issues of existing recommendation systems. The following are some suggestions of the future work for this domain problem.

1. It would be of interest to attempt to add more influencing factors in the processing stage. Relative location and abstract location of the user can be included for making recommendations.
2. The research direction can be extended to focus more on revisiting mobile content users. Establishing an individual prediction model based on content-based analysis and a user's preferences should be addressed.
3. Further study of data mining techniques for providing relevant information for a mobile content user can be investigated more. Other pre-processing and post-processing stages using feature selection techniques can be added on.
4. The rule pruning techniques can be included to reduce the number of rules according to a user profile and a user's preferences.

List of References

[1] L. Geddes, "Rat in your cellphone," *New Scientist*, pp. 34-37, Oct. 2009.

[2] S. J. H. Yang, J. Zhang, N. W. Y. Shao, and R. C. S. Chen, "A Service Supporting Universal Access to Mobile Internet with Unit of Information-Based Intelligent Content Adaptation," presented at the *IEEE International Conference on Web Services (ICWS 2007)*, 2007.

[3] B. Smyth and P. Cotter, "Intelligent Navigation for Mobile Internet Portals," in *IJCAI-03 Workshop on Artificial Intelligence, Information Access and Mobile Computing. The 18th International Joint Conference on Artificial Intelligence*, Acapulco, Mexico, 2003.

[4] D. Swallow, M. Blythe, and P. Wright, "Grounding Experience: Relating Theory and Method to Evaluate the User Experience of Smartphones," presented at *the EACE '05: Proceedings of the 2005 Annual Conference on European Association of Cognitive Ergonomics*, 2005.

[5] J. J. Cole, M. J. Gray, J. W. Lloyd, and K. S. Ng, "Personalisation for User Agents," in *Proceedings of the fourth international joint conference on Autonomous Agents and Multiagent Systems (AAMAS '05)*, Utrecht, Netherlands, 2005, pp. 603-610.

[6] Q. H. Mahmoud, E. Al-Masri, and Z. Wang, "Design and Implementation of a Smart System for Personalization and Accurate Selection of Mobile Services," *Requirements Engineering*, vol. 12, no. 4, pp. 221-230, Oct. 2007.

[7] P. Chandrasekaran and A. Joshi, "MobileIQ: A Framework for Mobile Information Access," in *The Third International Conference on Mobile Data Management (MDM '02)*, Singapore, 2002.

[8] M. Eirinaki and M. Vazirgiannis, "Web Mining for Web Personalization," *ACM Transactions on Internet Technology*, vol. 3, no. 1, pp. 1-27, Feb. 2003.

[9] D. Zhang, "Web Content Adaptation for Mobile Handheld Devices," *Communications of the ACM*, vol. 50, no. 2, pp. 75-79, Feb. 2007.

[10] F. H. d. Olmo and E. Gaudioso, "Evaluation of Recommender Systems: A New Approach," *Expert Systems with Applications*, vol. 35, no. 3, pp. 790-804, Oct. 2008.

[11] R. v. Meteren and M. v. Someren, "Using Content-Based Filtering for Recommendation," in *MLnetECML2000 Workshop*, 2000.

[12] K. C. Lee and S. Kwon, "Online Shopping Recommendation Mechanism and Its Influence on Consumer Decisions and Behaviors: A causal map approach," *Expert Systems with Applications*, vol. 35, no. 4, pp. 1567-1574, Nov. 2008.

[13] *Amazon*. Available: www.amazon.com

[14] *MovieLens*. Available: www.movielens.org

[15] C. Y. Kim, J. K. Lee, Y. H. Cho, and D. H. Kim, "VISCORS: A Visual-content Recommender for the Mobile Web," *IEEE Intelligent Systems*, vol. 19, no. 6, p. 32, 2004.

[16] B. N. Miller, I. Albert, S. K. Kam, and J. A. Konstan, and J. Riedl, "MovieLens Unplugged: Experiences with a Recommender System on Four Mobile Device," in *Intelligent User Interfaces*, Miami, FL, USA, 2003.

[17] M.-H. Kuo, L.-C. Chen, and C.-W. Liang, "Building and Evaluating a Location-Based Service Recommendation System with a Preference Adjustment Mechanism," *Expert Systems with Applications*, vol. 36, no. 2, March, 2009.

[18] H. J. Lee and S. J. Park, "MONERS: A news recommender for the Mobile Web," *Expert Systems with Applications*, vol. 32, pp. 143-150, 2007.

[19] P. Melville, R. J. Mooney, and R. Nagarajan, "Content-Boosted Collaborative Filtering for Improved Recommendations," in *Eighteenth National Conference on Artificial Intelligence (AAAI-2002)*, Edmonton, Canada, 2002, pp. 187-192.

[20] A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock, "Methods and Metrics for Cold-start Recommendations," in *25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2002)*, Tampere, Finland, 2002, pp. 253-260.

[21] M. Wagner, W.-T. Balke, R. Hirschfeld, and W. Kellerer, "A Roadmap to Advanced Personalization of Mobile Services," in *ODBASE, CoopIS 2002*, Irvine, CA, USA, 2002.

[22] W. P. Lee, "Deploying Personalized Mobile Service in an Agent-based Environment," *Expert Systems with Applications*, vol. 32, no. 4, pp. 1197-1207, May 2007.

[23] A. Ning, X. Guo, J. Cheng, and F. Meng, "Mobile Internet Services Personalization Customization via Mobile Portals," in *International Conference on Web Information Systems and Mining (WISM 2009)*, Shanghai, China, 2009.

[24] A. Kobsa, "Generic User Modeling Systems," *User Modeling and User-Adapted Interaction*, vol. 11, no. 1-2, pp. 49-63, 2001.

[25] N. Weißenberg, A. Voisard, and R. Gartmann, "Using Ontologies in Personalized Mobile Applications," in *Proceedings of the 12th Annual ACM International Workshop on Geographic Information Systems (GIS '04)*, Washington, DC, USA, 2004.

[26] J. Gemmell, G. Bell, and R. Leuder, "MyLifeBits: A Personal Database for Everything," *Communications of the ACM*, vol. 49, no. 1, pp. 88-95, Jan. 2006.

[27] M. Castells, M. Fernandez-Ardevol, J. L. Qiu, and A. Sey, *Mobile Communication and Society: A Global Perspective*. Cambridge, Massachusetts: The MIT Press, 2007.

[28] C. Seitz, M. Berger, and B. Bauer, "Towards a General Approach to Mobile Profile Based Distributed Grouping," *Personal and Ubiquitous Computing*, vol. 9, no. 2, pp. 90-99, March 2005.

[29] H. Jeon, T. Kim, and J. Choi, "Mobile Semantic Search using Personal Preference Filtering," presented at *the Fourth Internation Conference on Networked Computing and Advanced Information Management*, Gyeongju, Korea, 2008.

[30] J.-y. Hong, E.-h. Suh, and S.-j. Kim, "Context-aware Systems: A Literature Review and Classification," *Expert Systems with Applications*, vol. 36, no. 4, May 2009.

[31] G. Chen and D. Kotz, "A Survey of Context-Aware Mobile Computing Research," Department of Computer Science, Dartmouth College, Technical Report TR2000-381, November 2000.

[32] P. E. Román, G. L'Huillier, and J. D. Velásquez, "Web Usage Mining," in *Advanced Techniques in Web Intelligence - 1*, J. D. Velásquez and L. C. Jain, Eds., Berlin Heidelberg: Springer-Verlag, 2010, pp. 143-165.

[33] D.-R. Liu, C.-H. Lai, and W.-J. Lee, "A Hybrid of Sequential Rules and Collaborative Filtering for Product Recommendation," *Information Sciences*, vol. 179, no. 20, pp. 3505-3519, Sep. 2009.

[34] A. Montgomery and M. D. Smith, "Prospects for Personalization on the Internet," *Interactive Marketing*, vol. 23, no. 2, pp. 130-137, May 2009.

[35] I. Jorstad, D. V. Thanh, and S. Dustdar, "Personalisation of Future Mobile Services," in *9th International Conference on Intelligence in Service Delivery*, Bordeaux, France, 2004.

[36] C. Panayiotou and G. Samaras, "mPERSONA: Personalized Portals for the Wireless User: An Agent Approach," *Mobile Networks and Applications*, vol. 9, no. 6, pp. 663-677, Dec. 2004.

[37] J. Häkkilä and C. Chatfield, "Personal Customisation of Mobile Phones: A Case Study," presented at *Proceedings of the 4th Nordic Conference on Human-computer Interaction: Changing Roles (NordiCHI '06)*, Oslo, Norway, 2006.

[38] A. Oulasvirta and J. Blom, "Motivations in Personalisation Behaviour," *Interacting with Computers*, vol. 20, no. 1, pp. 1-16, Jan. 2007.
[39] L. Ventä, M. Isomursu, A. Ahtinen, and S. Ramiah, "'My Phone Is a Part of My Soul'--How People Bond with Their Mobile Phones," in *the Second International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM '08)*, Valencia, Spain, 2008.
[40] I. Cingil, A. Dogac, and A. Azgin, "A Broader Approach to Personalization," *Communications of the ACM*, vol. 43, no. 8, pp. 136-141, Aug. 2000.
[41] N. Arora, X. Dreze, A. Ghose, J. D. Hess, R. Iyengar, B. Jing, Y. Joshi, V. Kumar, N. Lurie, S. Neslin, S. Sajeesh, M. Su, N. Syam, J. Thomas, and Z. J. Zhang, "Putting One-to-one Marketing to Work: Personalization, Customization, and Choice," *Marketing Letters*, vol. 19, pp. 305-321, 2008.
[42] S. Perugini and N. Ramakrishnan, "Personalizing Interactions with Information Systems," in *Advances in Computers*, N. Zelkowitz, Ed., Academic Press, 2003, pp. 323-382.
[43] D. Zhang, G. Karabatis, Z. Chen, B. Adipat, L. Dai, Z. Zhang, and Y. Wang, "Personalization and Visualization on Handheld Devices," presented at *Proceedings of the 2006 ACM Symposium on Applied Computing (SAC '06)*, Dijon, France, 2006.
[44] Y. Borodin, J. Mahmud, and I. V. Ramakrishnan, "Context Browsing with Mobiles - When Less is More," in *Proceedings of the 5th International Conference on Mobile Systems, Applications and Services (MobiSys '07)*, San Juan, Puerto Rico, USA., 2007.
[45] S. Burigat, L. Chittaro, and S. Gabrielli, "Navigation Techiniques for Small-screen Devices: An Evaluation on Maps and Web Pages," *International Journal of Human-Computer Studies*, vol. 66, no. 2, pp. 78-97, Feb. 2008.
[46] X. Yin, W. S. Lee, and Z. Tan, "Personalization of Web Content for Wireless Mobile Device," in *IEEE Wireless Communications and Networking Conference*, Atlanta, GA, USA, 2004.
[47] M. Halvey, M. T. Keane, and B. Smyth, "Predicting Navigation Patterns on the Mobile-Internet Using Time of the Week," in *the 14th International Conference on World Wide Web (WWW '05)*, Chiba, Japan, 2005.
[48] A. Göker, S. Watt, H. I. Myrhaug, N. Whitehead, M. Yakici, R. Bierig, S. K. Nuti, and H. Cumming, "An Ambient, Personalised, and Context-Sensitive Information System for Mobile Users," presented at *the 2nd European Symposium on Ambient Intelligence (EUSAI 2004)*, Eindhoven, the Netherlands, 2004.
[49] A. Zipf and M. Jöst, "Implementing Adaptive Mobile GI Services Based on Ontologies: Examples from Pedestrian Navigation Support," *Computers, Environment and Urban Systems*, vol. 30, pp. 784-798, 2006.
[50] C. Feijóo, I. Maghiros, F. Abadie, and J.-L. Gómez-Barroso, "Exploring a Heterogeneous and Fragmented Digital Ecosystem: Mobile Content," *Telematics and Informatics*, vol. 26, no. 3, pp. 282-292, Aug. 2009.
[51] U. Shardanand and P. Maes, "Social Information Filtering: Algorithms for Automating 'Word of Mouth'," in *CHI '95 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Denver, Colorado, USA, 1995, pp. 210-217.
[52] M.-H. Park, J.-H. Hong, and S.-B. Cho, "Location-Based Recommendation System Using Bayesian User's Preference Model in Mobile Devices," *Ubiquitous Intelligence and Computing Lecture Notes in Computer Science*, vol. 4611, pp. 1130-1139, 2007.
[53] A. Y.-A. Chen and D. McLeod, "Collaborative Filtering for Information Recommendation Systems," in *Encyclopedia of E-Commerce, E-Government, and Mobile Commerce*. vol. 1, M. Khosrow-Pour, Ed., Idea Group Publishing 2006, pp. 118-123.

[54] Y. Yu, Z. Shanfeng, and C. Xinmeng, "Collaborative Filtering Algorithm Based on Kendall Correlation in Recommender Systems," *Wuhan University Journal of Natural Sciences,* vol. 11, no. 5, pp. 1086-1090, 2006.

[55] D.-R. Liu and C.-H. Liou, "Mobile Commerce Product Recommendations Based on Hybrid Multiple Channels," *Electronic Commerce Research and Applications,* vol. 10, no. 1, pp. 94-104, 2011.

[56] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based Collaborative Filtering Recommendation Algorithms," in *Proceedings of the 10th International Conference on World Wide Web (WWW '01),* Hong Kong, 2001.

[57] J. Ji, C. Liu, J. Yan, and N. Zhong, "Bayesian Networks Structure Learning and Its Application to Personalized Recommendation in a B2C Portal," in *IEEE/WIC/ACM International Conference on Web Intelligence 2004 (WI' 04)*, Beijing, China, 2004.

[58] S. Liao, T. Zou, and H. Chang, "An Association Rules and Sequential Rules Based Recommendation System," in *4th International Conference on Wireless Communications, Networking and Mobile Computing, 2008, WiCOM '08,* Dalian, China, 2008, pp. 1-4.

[59] D.-H. Kim, V. Atluri, I. Im, M. Bieber, N. Adam, and Y. Yesha, "A Clickstream-based Collaborative Filtering Personalization Model: Towards a Better Performance," in *Proceedings of the 6th Annual ACM International Workshop on Web Information and Data Management (WIDM '04)*, Washington, DC, USA., 2004.

[60] M. Papagelis and D. Plexousakis, "Qualitative Analysis of User-based and Item-based Prediction Algorithms for Recommendation Agents," *Engineering Applications of Artificial Intelligence,* vol. 18, pp. 781-789, 2005.

[61] G. Linden, B. Smith, and J. York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," *IEEE Internet Computing,* vol. 7, no. 1, pp. 76-80, Jan/Feb. 2003.

[62] J. Zhang, Z. Lin, B. Xiao, and C. Zhang, "An Optimized Item-based Collaborative Filtering Recommendation Algorithm," in *2009 IEEE International Conference on Network Infrastructure and Digital Content*, Beijing, China, 2009.

[63] D. Lemire and S. McGrath, "Implementing a Rating-Based Item-to-Item Recommender System in PHP/SQL," Université Du Québec Technical Report 2005.

[64] H. Liang, Y. Xu, Y. Li, and R. Nayak, "Collaborative Filtering Recommender Systems Using Tag Information," presented at *the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology,* Sydney, Australia, 2008.

[65] Y.-n. Chen and M. Yu, "A Hybrid Collaborative Filtering Algorithm Based on User-Item," presented at *the International Conference on Computational and Information Sciences*, Chengdu, Sichuan, China, 2010.

[66] S. K. Lee, Y. H. Cho, and S. H. Kim, "Collaborative Filtering with Ordinal Scale-based Implicit Ratings for Mobile Music Recommendations," *Information Sciences,* vol. 180, pp. 2142-2155, 2010.

[67] M. J. Pazzani, "A Framework for Collaborative, Content-Based and Demographic Filtering," *Artificial Intelligence Review,* vol. 13, no. 5-6, pp. 398-408, Dec. 1999.

[68] T. W. Yan and H. Garcia-Molina, "SIFT: A Tool for Wide-area Information Dissemination," in *Proceedings of the USENIX 1995 Technical Conference (TCON'95)*, 1995, pp. 177-186.

[69] G. Xu, Y. Zhang, and L. Li, *Web Mining and Social Networking: Techniques and Applications*, Springer, 2010.

[70] P.-H. Chiu, G. Y.-M. Kao, and C.-C. Lo, "Personalized Blog Content Recommender System for Mobile Phone Users," *Internation Journal of Human-Computer Studies,* vol. 68, no. 8, pp. 496-507, 2010.

[71] I. Bose and C. Xi, "Exploring Business Opportunities from Mobile Service Data of Customers Using Inter-cluster Analysis," in *The IEEE International Workshop on Data Mining for Design and Marketing*, Hong Kong, 2006, pp. 1-5.

[72] T. Yamakami, "Toward Understanding the Mobile Internet User Behavior: A Methodology for User Clustering with Aging Analysis," presented at *the Fourth International Conference on Parallel and Distributed Computing, Applications and Technologies*, Chengdu, China, 2003.

[73] M. N. Cortimiglia, A. Ghezzi, and F. Renga, "Mobile Applications and Their Delivery Platforms," *IT Professional*, vol. 13, no. 5, pp. 51-56, 2011.

[74] F. Zhiyuan, Z. Jingchun, W. Lei, Z. Yu, and M. Minghui, "MCTSE: A Commerce Topic Faced P2P Search Engine for Mobile Devices," in *IEEE International Conference on e-Business Engineering*, Hong Kong, China, 2007, pp. 485-491.

[75] K. Csorba and I. Vajk, "Taxonomic Support for Document Classification in Mobile Device Environment," in *Conference on Human System Interactions*, Krakow, Poland, 2008, pp. 674-679.

[76] G. Shani, L. Rokach, A. Meisles, L. Naamani, N. Piratla, and D. Ben-Shimon, "Establishing User Profiles in the MediaScout Recommender System," in *IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2007)*, Honolulu, Hawaii, USA, 2007.

[77] J. Srivastava, R. Cooley, M. Deshpande, and P.-N. Tan, "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data," *SIGKDD Explorations*, vol. 1, no. 2, pp. 12-23, Jan. 2000.

[78] B. Liu, *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. New York: Springer, 2007.

[79] M. H. Dunham, *Data Mining: Introductory and Advanced Topics*. New Jersey: Pearson Education Inc., 2003.

[80] E. Turban, J. E. Aronson, T.-P. Liang, and R. Sharda, *Decision Support and Business Intelligence Systems*, 8th ed. New Jersey: Pearson Prentice Hall, 2006.

[81] J. Fink, J. Koenemann, S. Noller, and I. Schwab, "Putting Personalisation into Practice," *Communications of the ACM*, vol. 45, no. 5, pp. 41-42, May 2002.

[82] B. Smyth and P. Cotter, "Personalized Adaptive Navigation for Mobile Portals," in *15th European Conference on Artificial Intelligent—Prestigious Application of Intelligent Systems*, Lyons, France, 2002.

[83] Y. Zhang and J. r. Jiao, "An Associative Classification-based Recommendation System for Personalization in B2C E-commerce Applications," *Expert Systems with Applications*, vol. 33, no. 2, pp. 357-367, Aug. 2007.

[84] I. Bose and R. K. Mahapatra, "Business Data Mining—A Machine Learning Perspective," *Information and Management*, vol. 39, no. 3, pp. 211-225, 2001.

[85] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. McLachlan, A. Ng, B. Liu, P. Yu, Z.-H. Zhou, M. Steinbach, D. Hand, and D. Steinberg, "Top 10 Algorithms in Data Mining," *Knowledge and Information Systems*, vol. 14, no. 1, pp. 1-37, 2008.

[86] A. K. Jain, "Data Clustering: 50 Years Beyond K-Means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651-666, 2009.

[87] W. Paireekreng and K. W. Wong, "Mobile Content Personalisation Using Intelligent User Profile Approach," in *The 3rd International Conference on Knowledge Discovery and Data Mining (WKDD 2010)*, Phuket, Thailand, 2010.

[88] W. Paireekreng and K. W. Wong, "Client-side Mobile User Profile for Content Management Using Data Mining Techniques," in *8th International Symposium on Natural Language Processing*, Bangkok, Thailand, 2009.

[89] S. Okazaki, "What Do We Know About Mobile Internet Adopters? A Cluster Analysis," *Information and Management*, vol. 43, no. 2, pp. 127-141, March 2006.

[90] T. Yamakami, "A User-Perceived Freshness Clustering Method to Identify Three Subgroups in Mobile Internet Users," presented at *the 2nd International Conference on Multimedia and Ubiquitous Engineering*, Busan, Korea, 2008.
[91] T. Yamakami, "Exploratory Day-scale Behavior Assumption-Based User Clustering with the Mobile Clickstream," presented at *the Eighth International Conference on Parallel and Distributed Computing, Application and Technologies*, Adelaide, Australia, 2007.
[92] S. Y. Sohn and Y. Kim, "Searching Customer Patterns of Mobile Service Using Clustering and Quantitative Association Rule," *Expert Systems with Applications,* vol. 34, no. 2, pp. 1070-1077, Feb. 2008.
[93] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the Number of Clusters in a Data Set via the Gap Statistic," *Journal of the Royal Statistical Society*, vol. 63, pp. 411-423, 2001.
[94] M. Muhr and M. Granitzer, "Automatic Cluster Number Selection using a Split and Merge K-Means Approach," in *20th International Workshop on Database and Expert Systems Application*, Linz, Austria, 2009, pp. 363-367.
[95] S. Salvador and P. Chan, "Determining the Number of Clusters/Segments in Hierarchical Clustering/Segmentation Algorithms," in *16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'04)*, Boca Raton, Florida, 2004.
[96] J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles*. Massachusetts: Addison-Wesley, 1974.
[97] G. Shmueli, N. R. Patel, and P. C. Bruce, *Data Mining for Business Intelligence*. New Jersey: Wiley & Sons, 2010.
[98] P. Nurmi, M. Hassinen, and K. C. Lee, "A Comparative Analysis of Personalization Techniques for a Mobile Application," presented at *the 21st International Conference on Advanced Information Networking and Applications Workshops (AINAW '07)*, 2007.
[99] A. Cufoglu, M. Lohi, and K. Madani, "A Comparative Study of Selected Classifiers with Classification Accuracy in User Profiling," presented at *the 2009 WRI World Congress on Computer Science and Information Engineering*, Los Angeles, California, USA 2009.
[100] S. B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques," *Informatica,* vol. 31, pp. 249-268, 2007.
[101] R. Agrawal, T. Imieliński, and A. Swami, "Mining Association Rules between Sets of Items in Large Databases," in *ACM-SIGMOD International Conference on Management of Data*, 1993, pp. 207-216.
[102] G. Shmueli, N. R. Patel, and P. C. Bruce, *Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Excel with XLMiner,* New Jersey: Wiley, 2007.
[103] M. N. Kumar and B. E. Reddy, "Improved Classification Association Rule Mining," presented at *the International Conference on Intelligent Agent & Multi-Agent Systems, 2009 (IAMA 2009),* Chennai, India, 2009.
[104] H. Ye, "A Personalized Collaborative Filtering Recommendation Using Association Rules Mining and Self-Organizing Map," *Journal of Software*, vol. 6, no. 4, pp. 732-739, April 2011.
[105] B. Mobasher, H. Dai, T. Luo, and M. Nakagawa, "Effective Personalization Based on Association Rule Discovery from Web Usage Data," in *Proceedings of the ACM Workshop on Web Information and Data Management (WIDM '01)*, Atlanta, GA, USA, 2001.
[106] M. Hillborg, *Wireless XML Developer's Guide*. California: McGraw-Hill/Osbourne, 2002.

[107] M. Sasajima, Y. Kitamura, T. Naganuma, S. Kurakake, and R. Mizoguchi, "Toward Task Ontology-based Modeling for Mobile Phone Users' Activity," in *4th International Semantic Web Conference (ISWC 2005)*, Galway, Ireland, 2005.
[108] G. Buchanan, S. Farrant, M. Jones, H. Thimbleby, G. Marsden, and M. Pazzani, "Improving Mobile Internet Usability," in *Proceedings of the 10th International Conference on World Wide Web (WWW '01)*, Hong Kong, China, 2001.
[109] J. T. S. Quah and V. L. H. Seet, "Adaptive WAP Portals," *Electronic Commerce Research and Applications*, vol. 7, no. 4, pp. 377-385, 2009.
[110] D. K. Yun, K. Y. Kim, and H. S. Ko, "Customer Expectation Level in Mobile Data Services," in *Proceedings of the 7th International Conference on Human Computer Interaction with Mobile Devices & Services (MobileHCI '05)*, Salzburg, Austria, 2005.
[111] M. Halvey, M. T. Keane, and B. Smyth, "Time-based Segmentation of Log Data for User Navigation Prediction in Personalization," in *IEEE/WIC/ACM International Conference on Web Intelligence (WI' 05)*, France, 2008.
[112] C. R. Anderson, P. Domingos, and D. S. Weld, "Adaptive Web Navigation for Wireless Devices," in *Seventeenth International Joint Conference on Artificial Intelligence (IJCAI-01)*, Seattle, Washington, USA, 2001.
[113] A. A. Ghorbani and X. Xu, "A Fuzzy Markov Model Approach for Predicting User Navigation," in *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (WI '07)*, 2007.
[114] B. Piwowarski and H. Zaragoza, "Predictive User Click Models Based on Click-through History," in *Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management (CIKM'07)*, Lisboa, Portugal, 2007.
[115] D. J. Xu, S. S. Liao, and Q. Li, "Combining Empirical Experimentation and Modeling Techniques: A Design Research Approach for Personalized Mobile Advertising Applications," *Decision Support Systems*, vol. 44, no. 3, pp. 710-724, Feb. 2008.
[116] H. Liu and V. Kešelj, "Combined Mining of Web Server Logs and Web Contents for Classifying User Navigation Patterns and Predicting Users's Future Requests," *Data & Knowledge Engineering*, vol. 61, no. 2, pp. 304-330, May 2007.
[117] M. Levene and R. Wheeldon, "A Web Site Navigation Engine," in *Proceedings of the 10th International Conference on World Wide Web (WWW'10)*, Hong Kong, China, 2001.
[118] J. McMullan and I. Richardson, "The Mobile Phone: A Hybrid Multi-platform Medium," in *Proceedings of the 3rd Australasian Conference on Interactive Environment (IE'06)*, 2006.
[119] C. Albanesius. (2011, May 2, 2011) Google Sued Over Android Location Services. *PC Magazine*.
[120] A. Göker and H. I. Myrhaug, "User Context and Personalisation," in *Proceedings of the 6th European Conference on Case Based Reasoning Workshops (ECCBR 2002)*, Aberdeen, Scotland, 2002.
[121] F.-H. Wang and H.-M. Shao, "Effective Personalized Recommendation Based on Time-Framed Navigation Clustering and Association Mining," *Expert Systems with Applications*, vol. 27, pp. 365-377, 2004.
[122] U. Gonzalez-Barron and F. Butler, "A Comparison of Seven Thresholding Techniques with the K-means Clustering Algorithm for Measurement of Bread-crumb Features by Digital Image Analysis," *Journal of Food Engineering*, vol. 74, pp. 268-278, 2006.
[123] S. Ray and R. H. Turi, "Determination of Number of Clusters in K-Means Clustering and Application in Colour Image Segmentation," in *the 4th International Conference on Advances in Pattern Recognition and Digital Techniques (ICAPRDT '99)*, Calcutta, India, 1999.

[124] G. Hussein, "Mobile Recommender System Analysis & Design," in *First International Conference on Networked Digital Technologies NDT '09*, Ostrava, The Czech Republic, 2009.
[125] D. S. Phatak and R. Mulvaney, "Clustering for Personalized Mobile Web Page," in *The IEEE International Conference on Fuzzy Systems*, 2002.
[126] W. Paireekreng, "Influence Factors of Mobile Content Personalization on Mobile Device User in Bangkok," Research report, 2007.
[127] M. Norusis, *IBM SPSS Statistics 19 Statistical Procedures Companion*. Addison Wesley, 2011.
[128] Z. Yun and F. Boqin, "Tag-based user modeling using formal concept analysis," presented at *the 8th IEEE International Conference on Computer and Information Technology (CIT 2008)* , Sydney, NSW, Australia, 2008.
[129] J. Teevan, S. T. Dumais, and E. Horvitz, "Personalizing Search via Automated Analysis of Interests and Activities," in *the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '05)*, Salvador, Brazil, 2005.
[130] P. Germanakos, N. Tsianos, Z. Lekkas, C. Mourlas, and G. Samaras, "Capturing Essential Intrinsic User Behaviour Values for the Design of Comprehensive Web-based Personalized Environments," *Computers in Human Behavior*, vol. 24, no. 4, pp. 1434-1451, July 2008.
[131] P. I. Hofgesang, "Web Personalisation Through Incremental Individual Profiling and Support-based User Segmentation," presented at *the IEEE/WIC/ACM International Conference on Web Intelligence (WI '07)*, 2007.
[132] V. M. García-Barrios, "User-centric Privacy Framework: Integrating Legal, Technological and Human Aspects into User-Adapting Systems," presented at *the International Conference on Computational Science and Engineering 2009 (CSE '09)*, Vancouver, BC, 2009.
[133] K. Partridge and P. Golle, "On Using Existing Time-use Study Data for Ubiquitous Computing Applications," in *10th International Conference on Ubiquitous Computing (UbiComp '08)*, Seoul, South Korea, 2008.

APPENDIX

MTCAR generated recommendation results for Dataset A

Item # User ID	1	2	3	4	5	6	7	8	9	10
1	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
2	IM14	IM311	IM31	IM312	IM315	IM11	IM21	IM33	IM23	IM22
3	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
4	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
5	IM14	IM312	IM313	IM31	IM326	IM315	IM32	IM316	IM327	IM311
6	IM31	IM312	IM313	IM326	IM315	IM14	IM311	IM327		
7	IM14	IM312	IM313	IM31	IM326	IM315	IM32	IM316	IM327	IM311
8	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
9	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
10	IM31	IM311	IM326	IM315	IM314	IM312	IM14			
11	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
12	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
13	IM14	IM311	IM31	IM312	IM313	IM23	IM32	IM326	IM11	IM315
14	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
15	IM14	IM311	IM31	IM312	IM315	IM11	IM21	IM33	IM23	IM22
16	IM14	IM311	IM31	IM312	IM315	IM11	IM21	IM33	IM23	IM22
17	IM14	IM311	IM31	IM312	IM313	IM23	IM32	IM326	IM11	IM315
18	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
19	IM312	IM313	IM31	IM326	IM315	IM14	IM311			
20	IM14	IM311	IM31	IM312	IM314	IM315	IM326	IM316	IM313	IM32
21	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
22	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
23	IM14	IM311	IM31	IM312	IM315	IM11	IM21	IM33	IM23	IM22
24	IM14	IM311	IM31	IM312	IM314	IM315	IM326	IM316	IM313	IM32
25	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
26	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
27	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
28	IM14	IM311	IM31	IM312	IM315	IM11	IM21	IM33	IM23	IM22
29	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
30	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
31	IM14	IM311	IM31	IM312	IM315	IM11	IM21	IM33	IM23	IM22
32	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
33	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
34	IM312	IM313	IM31	IM326	IM315	IM14	IM311			
35	IM14	IM311	IM31	IM312	IM315	IM11	IM21	IM33	IM23	IM22
36	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
37	IM14	IM311	IM31	IM312	IM313	IM23	IM32	IM326	IM11	IM315
38	IM312	IM313	IM31	IM326	IM315	IM14	IM311			
39	IM14	IM311	IM31	IM312	IM315	IM11	IM21	IM33	IM23	IM22
40	IM14	IM311	IM31	IM312	IM313	IM23	IM32	IM326	IM11	IM315
41	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
42	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
43	IM31	IM312	IM313	IM326	IM315	IM14	IM311	IM327		
44	IM14	IM311	IM31	IM312	IM315	IM11	IM21	IM33	IM23	IM22
45	IM14	IM311	IM31	IM312	IM314	IM315	IM326	IM316	IM313	IM32
46	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
47	IM14	IM311	IM31	IM312	IM315	IM11	IM21	IM33	IM23	IM22
48	IM14	IM311	IM31	IM312	IM315	IM11	IM21	IM33	IM23	IM22
49	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
50	IM14	IM312	IM313	IM31	IM326	IM315	IM32	IM316	IM327	IM311

Item # User ID	1	2	3	4	5	6	7	8	9	10
51	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
52	IM14	IM311	IM31	IM312	IM315	IM11	IM21	IM33	IM23	IM22
53	IM14	IM311	IM31	IM312	IM314	IM315	IM326	IM316	IM313	IM32
54	IM14	IM311	IM31	IM312	IM315	IM11	IM21	IM33	IM23	IM22
55	IM312	IM313	IM31	IM326	IM315	IM14	IM311			
56	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
57	IM14	IM312	IM313	IM31	IM326	IM315	IM32	IM316	IM327	IM311
58	IM311	IM312	IM313	IM31	IM23	IM32	IM326	IM11	IM14	IM315
59	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
60	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
61	IM14	IM312	IM313	IM31	IM326	IM315	IM32	IM316	IM327	IM311
62	IM14	IM311	IM31	IM312	IM314	IM315	IM326	IM316	IM313	IM32
63	IM14	IM311	IM31	IM312	IM315	IM11	IM21	IM33	IM23	IM22
64	IM14	IM311	IM31	IM312	IM315	IM11	IM21	IM33	IM23	IM22
65	IM14	IM312	IM313	IM31	IM326	IM315	IM32	IM316	IM327	IM311
66	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
67	IM14	IM312	IM313	IM31	IM326	IM315	IM32	IM316	IM327	IM311
68	IM312	IM313	IM31	IM326	IM315	IM14	IM311			
69	IM14	IM311	IM31	IM312	IM315	IM11	IM21	IM33	IM23	IM22
70	IM14	IM311	IM31	IM312	IM314	IM315	IM326	IM316	IM313	IM32
71	IM14	IM311	IM31	IM312	IM315	IM11	IM21	IM33	IM23	IM22
72	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
73	IM312	IM313	IM31	IM14	IM326	IM315	IM32	IM316	IM327	IM311
74	IM14	IM312	IM31	IM313	IM315	IM32	IM326	IM33	IM11	IM316
75	IM311	IM31	IM312	IM313	IM23	IM32	IM326	IM11	IM14	
76	IM14	IM312	IM313	IM31	IM326	IM315	IM32	IM316	IM327	IM311
77	IM14	IM311	IM31	IM312	IM314	IM315	IM326	IM316	IM313	IM32
78	IM14	IM311	IM31	IM312	IM314	IM315	IM326	IM316	IM313	IM32
79	IM14	IM311	IM31	IM312	IM315	IM11	IM21	IM33	IM23	IM22
80	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
81	IM14	IM312	IM31	IM313	IM315	IM32	IM326	IM33	IM11	IM316
82	IM14	IM311	IM31	IM312	IM314	IM315	IM326	IM316	IM313	IM32
83	IM14	IM311	IM31	IM312	IM315	IM11	IM21	IM33	IM23	IM22
84	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
85	IM14	IM311	IM31	IM312	IM314	IM315	IM326	IM316	IM313	IM32
86	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
87	IM14	IM311	IM31	IM312	IM314	IM315	IM326	IM316	IM313	IM32
88	IM14	IM312	IM31	IM313	IM315	IM32	IM326	IM33	IM11	IM316
89	IM31	IM312	IM313	IM326	IM315	IM14	IM311	IM327		
90	IM14	IM311	IM31	IM312	IM315	IM11	IM21	IM33	IM23	IM22
91	IM311	IM31	IM312	IM313	IM23	IM32	IM326	IM11	IM14	
92	IM14	IM311	IM31	IM312	IM315	IM313	IM11	IM21	IM22	IM316
93	IM14	IM311	IM31	IM312	IM315	IM11	IM21	IM33	IM23	IM22
94	IM31	IM312	IM313	IM326	IM315	IM14	IM311	IM327		
95	IM14	IM312	IM313	IM31	IM326	IM315	IM32	IM316	IM327	IM311
96	IM14	IM311	IM31	IM312	IM315	IM11	IM21	IM33	IM23	IM22
97	IM31	IM312	IM313	IM326	IM315	IM14	IM311	IM327		
98	IM14	IM312	IM313	IM31	IM326	IM315	IM32	IM316	IM327	IM311
99	IM311	IM312	IM313	IM31	IM23	IM32	IM326	IM11	IM14	IM315
100	IM14	IM311	IM31	IM312	IM315	IM11	IM21	IM33	IM23	IM22

* Please note that 'IM' stands for item number

MTCAR generated recommendation results for Dataset B

Item # User ID	1	2	3	4	5	6	7	8	9	10
1	IM311	IM14	IM31	IM312	IM326	IM327	IM315	IM313	IM316	IM32
2	IM14	IM311	IM11	IM21	IM312	IM31	IM22	IM315	IM313	IM316
3	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
4	IM14	IM311	IM31	IM312	IM315	IM11	IM33	IM21	IM23	IM32
5	IM311	IM14	IM31	IM312	IM326	IM327	IM315	IM313	IM316	IM32
6	IM14	IM311	IM31	IM312	IM315	IM11	IM33	IM21	IM23	IM32
7	IM311	IM14	IM31	IM312	IM326	IM327	IM315	IM313	IM316	IM32
8	IM14	IM311	IM31	IM312	IM315	IM11	IM33	IM21	IM23	IM32
9	IM311	IM14	IM31	IM312	IM326	IM327	IM315	IM313	IM316	IM32
10	IM14	IM311	IM11	IM312	IM31	IM21	IM22	IM315	IM313	IM316
11	IM311	IM31	IM312	IM313	IM326	IM315	IM14			
12	IM311	IM31	IM312							
13	IM14	IM311	IM11	IM312	IM31	IM21	IM22	IM315	IM313	IM316
14	IM14	IM311	IM11	IM21	IM312	IM31	IM22	IM315	IM313	IM316
15	IM14	IM311	IM31	IM312	IM315	IM11	IM33	IM21	IM23	IM32
16	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
17	IM14	IM311	IM31	IM312	IM315	IM11	IM33	IM21	IM23	IM32
18	IM14	IM311	IM11	IM21	IM312	IM31	IM22	IM315	IM313	IM316
19	IM14	IM311	IM31	IM312	IM315	IM11	IM33	IM21	IM23	IM32
20	IM311	IM31	IM312	IM326	IM14	IM313	IM316	IM315	IM32	IM314
21	IM14	IM311	IM11	IM21	IM312	IM31	IM22	IM315	IM313	IM316
22	IM14	IM311	IM31	IM312	IM315	IM11	IM33	IM21	IM23	IM32
23	IM313	IM311	IM312	IM326	IM315	IM14	IM31			
24	IM312	IM326	IM31	IM14	IM313	IM316	IM315	IM32	IM314	IM327
25	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
26	IM14	IM311	IM31	IM312	IM315	IM11	IM33	IM21	IM23	IM32
27	IM14	IM311	IM11	IM21	IM312	IM31	IM22	IM315	IM313	IM316
28	IM14	IM311	IM11	IM312	IM31	IM21	IM22	IM315	IM313	IM316
29	IM312	IM326	IM31	IM14	IM313	IM316	IM315	IM32	IM314	IM327
30	IM14	IM311	IM11	IM312	IM31	IM21	IM22	IM315	IM313	IM316
31	IM14	IM311	IM11	IM312	IM31	IM21	IM22	IM315	IM313	IM316
32	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
33	IM312	IM326	IM31	IM14	IM313	IM316	IM315	IM32	IM314	IM327
34	IM311	IM14	IM31	IM312	IM326	IM327	IM315	IM313	IM316	IM32
35	IM14	IM311	IM31	IM312	IM315	IM11	IM33	IM21	IM23	IM32
36	IM14	IM311	IM11	IM312	IM31	IM21	IM22	IM315	IM313	IM316
37	IM14	IM311	IM11	IM312	IM31	IM21	IM22	IM315	IM313	IM316
38	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
39	IM312	IM326	IM31	IM14	IM313	IM316	IM315	IM32	IM314	IM327
40	IM312	IM326	IM31	IM14	IM313	IM316	IM315	IM32	IM314	IM327
41	IM311	IM31	IM312	IM313	IM326	IM315	IM14			
42	IM313	IM311	IM312	IM326	IM315	IM14	IM31			
43	IM14	IM311	IM31	IM312	IM315	IM11	IM33	IM21	IM23	IM32
44	IM14	IM311	IM11	IM312	IM31	IM21	IM22	IM315	IM313	IM316
45	IM14	IM311	IM11	IM21	IM312	IM31	IM22	IM315	IM313	IM316
46	IM14	IM311	IM31	IM312	IM315	IM11	IM33	IM21	IM23	IM32
47	IM14	IM311	IM11	IM312	IM31	IM21	IM22	IM315	IM313	IM316
48	IM14	IM311	IM11	IM312	IM31	IM21	IM22	IM315	IM313	IM316
49	IM312	IM326	IM31	IM14	IM313	IM316	IM315	IM32	IM314	IM327
50	IM14	IM311	IM11	IM312	IM31	IM21	IM22	IM315	IM313	IM316

Item # User ID	1	2	3	4	5	6	7	8	9	10
51	IM311	IM31	IM312	IM313	IM326	IM315	IM14			
52	IM14	IM311	IM31	IM312	IM315	IM11	IM33	IM21	IM23	IM32
53	IM311	IM14	IM31	IM312	IM326	IM327	IM315	IM313	IM316	IM32
54	IM312	IM326	IM31	IM14	IM313	IM316	IM315	IM32	IM314	IM327
55	IM311	IM31	IM312	IM313	IM326	IM315	IM14			
56	IM311	IM14	IM31	IM312	IM326	IM327	IM315	IM313	IM316	IM32
57	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
58	IM14	IM311	IM31	IM312	IM315	IM11	IM33	IM21	IM23	IM32
59	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
60	IM14	IM311	IM11	IM21	IM312	IM31	IM22	IM315	IM313	IM316
61	IM14	IM311	IM31	IM312	IM315	IM11	IM33	IM21	IM23	IM32
62	IM312	IM326	IM31	IM14	IM313	IM316	IM315	IM32	IM314	IM327
63	IM14	IM311	IM31	IM312	IM315	IM11	IM33	IM21	IM23	IM32
64	IM313	IM311	IM312	IM326	IM315	IM14	IM31			
65	IM312	IM326	IM31	IM14	IM313	IM316	IM315	IM32	IM314	IM327
66	IM14	IM311	IM11	IM21	IM312	IM31	IM22	IM315	IM313	IM316
67	IM311	IM31	IM312							
68	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
69	IM311	IM14	IM31	IM312	IM326	IM327	IM315	IM313	IM316	IM32
70	IM311	IM14	IM31	IM312	IM326	IM327	IM315	IM313	IM316	IM32
71	IM14	IM311	IM31	IM312	IM315	IM11	IM33	IM21	IM23	IM32
72	IM14	IM311	IM11	IM312	IM31	IM21	IM22	IM315	IM313	IM316
73	IM14	IM311	IM31	IM312	IM315	IM11	IM33	IM21	IM23	IM32
74	IM14	IM311	IM31	IM312	IM315	IM11	IM33	IM21	IM23	IM32
75	IM311	IM31	IM312	IM313	IM326	IM315	IM14			
76	IM313	IM311	IM312	IM326	IM315	IM14	IM31			
77	IM14	IM311	IM31	IM312	IM315	IM11	IM33	IM21	IM23	IM32
78	IM311	IM14	IM31	IM312	IM326	IM327	IM315	IM313	IM316	IM32
79	IM14	IM311	IM11	IM312	IM31	IM21	IM22	IM315	IM313	IM316
80	IM14	IM311	IM31	IM312	IM315	IM11	IM33	IM21	IM23	IM32
81	IM14	IM311	IM31	IM312	IM315	IM11	IM33	IM21	IM23	IM32
82	IM14	IM311	IM11	IM312	IM31	IM21	IM22	IM315	IM313	IM316
83	IM312	IM31	IM14	IM313	IM32	IM311	IM23			
84	IM311	IM14	IM31	IM312	IM326	IM327	IM315	IM313	IM316	IM32
85	IM14	IM311	IM31	IM312	IM315	IM11	IM33	IM21	IM23	IM32
86	IM311	IM31	IM312	IM313	IM326	IM315	IM14			
87	IM14	IM311	IM11	IM312	IM31	IM21	IM22	IM315	IM313	IM316
88	IM311	IM14	IM31	IM312	IM326	IM327	IM315	IM313	IM316	IM32
89	IM14	IM311	IM11	IM21	IM312	IM31	IM22	IM315	IM313	IM316
90	IM313	IM311	IM312	IM326	IM315	IM14	IM31			
91	IM313	IM311	IM312	IM326	IM315	IM14	IM31			
92	IM312	IM313	IM31	IM326	IM315	IM14	IM311			
93	IM14	IM311	IM31	IM312	IM315	IM11	IM33	IM21	IM23	IM32
94	IM14	IM311	IM11	IM312	IM31	IM21	IM22	IM315	IM313	IM316
95	IM313	IM311	IM312	IM326	IM315	IM14	IM31			
96	IM14	IM311	IM11	IM21	IM312	IM31	IM22	IM315	IM313	IM316
97	IM14	IM311	IM11	IM21	IM312	IM31	IM22	IM315	IM313	IM316
98	IM312	IM326	IM31	IM14	IM313	IM316	IM315	IM32	IM314	IM327
99	IM312	IM31	IM14	IM313	IM32	IM311	IM23			
100	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23

* Please note that ‘IM’ stands for item number

MTCAR generated recommendation results for Dataset C

Item # User ID	1	2	3	4	5	6	7	8	9	10
1	IM312	IM31	IM311	IM11	IM33	IM21	IM23	IM32	IM12	IM326
2	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
3	IM311	IM31	IM312	IM11	IM21	IM22	IM315	IM14	IM313	IM316
4	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
5	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
6	IM14	IM311	IM31	IM312	IM326	IM327	IM315	IM313	IM316	IM32
7	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
8	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
9	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
10	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
11	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
12	IM14	IM311	IM31	IM312	IM326	IM327	IM315	IM313	IM316	IM32
13	IM311	IM31	IM312	IM11	IM21	IM22	IM315	IM14	IM313	IM316
14	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
15	IM14	IM311	IM31	IM312	IM326	IM327	IM315	IM313	IM316	IM32
16	IM311	IM31	IM312	IM11	IM21	IM22	IM315	IM14	IM313	IM316
17	IM311	IM31	IM312	IM11	IM21	IM22	IM315	IM14	IM313	IM316
18	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
19	IM311	IM31	IM312	IM11	IM21	IM22	IM315	IM14	IM313	IM316
20	IM14	IM311	IM31	IM312	IM326	IM327	IM315	IM313	IM316	IM32
21	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
22	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
23	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
24	IM31	IM311	IM312	IM326	IM14	IM313	IM315	IM314	IM316	IM327
25	IM14	IM311	IM31	IM313	IM326	IM312	IM315			
26	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
27	IM14	IM311	IM31	IM313	IM326	IM312	IM315			
28	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
29	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
30	IM14	IM311	IM31	IM312	IM326	IM327	IM315	IM313	IM316	IM32
31	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
32	IM313	IM312	IM31	IM326	IM315	IM14	IM311			
33	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
34	IM311	IM31	IM312	IM11	IM21	IM22	IM315	IM14	IM313	IM316
35	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
36	IM31	IM311	IM313	IM326	IM315	IM14				
37	IM31	IM311	IM312	IM326	IM14	IM313	IM315	IM314	IM316	IM327
38	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
39	IM14	IM311	IM31	IM313	IM326	IM312	IM315			
40	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
41	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
42	IM311	IM31	IM11	IM312	IM21	IM22				
43	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
44	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
45	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
46	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
47	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
48	IM14	IM311	IM31	IM312	IM326	IM327	IM315	IM313	IM316	IM32
49	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
50	IM31	IM311	IM313	IM326	IM315	IM14				

Item # User ID	1	2	3	4	5	6	7	8	9	10
51	IM312	IM31	IM311	IM11	IM33	IM21	IM23	IM32	IM12	IM326
52	IM14	IM311	IM31	IM312	IM326	IM327	IM315	IM313	IM316	IM32
53	IM14	IM311	IM31	IM313	IM326	IM312	IM315			
54	IM14	IM311	IM31	IM312	IM326	IM327	IM315	IM313	IM316	IM32
55	IM14	IM311	IM31	IM313	IM326	IM312	IM315			
56	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
57	IM311	IM31	IM312	IM11	IM21	IM22	IM315	IM14	IM313	IM316
58	IM311	IM31	IM11	IM312	IM21	IM22				
59	IM313	IM312	IM31	IM326	IM315	IM14	IM311			
60	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
61	IM313	IM312	IM31	IM326	IM315	IM14	IM311			
62	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
63	IM313	IM312	IM31	IM326	IM315	IM14	IM311			
64	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
65	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
66	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
67	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
68	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
69	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
70	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
71	IM311	IM31	IM312	IM11	IM21	IM22	IM315	IM14	IM313	IM316
72	IM311	IM31	IM312	IM11	IM21	IM22	IM315	IM14	IM313	IM316
73	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
74	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
75	IM14	IM311	IM31	IM312	IM326	IM327	IM315	IM313	IM316	IM32
76	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
77	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
78	IM14	IM311	IM31	IM312	IM326	IM327	IM315	IM313	IM316	IM32
79	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
80	IM311	IM31	IM312	IM11	IM21	IM22	IM315	IM14	IM313	IM316
81	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
82	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
83	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
84	IM31	IM311	IM313	IM326	IM315	IM14				
85	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
86	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
87	IM311	IM31	IM312	IM11	IM21	IM22	IM315	IM14	IM313	IM316
88	IM14	IM311	IM31	IM312	IM326	IM327	IM315	IM313	IM316	IM32
89	IM311	IM31	IM312	IM11	IM21	IM22	IM315	IM14	IM313	IM316
90	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
91	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
92	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
93	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
94	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
95	IM311	IM31	IM312	IM11	IM21	IM22	IM315	IM14	IM313	IM316
96	IM14	IM311	IM31	IM312	IM11	IM12	IM21	IM32	IM313	IM23
97	IM311	IM31	IM312	IM11	IM21	IM22	IM315	IM14	IM313	IM316
98	IM14	IM311	IM312	IM315	IM31	IM11	IM33	IM21	IM23	IM32
99	IM313	IM312	IM31	IM326	IM315	IM14	IM311			
100	IM31	IM311	IM312	IM326	IM14	IM313	IM315	IM314	IM316	IM327

* Please note that ‘IM’ stands for item number