A CONTEXT-AWARE APPLICATION RECOMMENDATION SYSTEM FOR MOBILE DEVICE USERS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

GAMZE BAYRAM

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
THE DEPARTMENT OF INFORMATION SYSTEMS

JANUARY 2015

A CONTEXT-AWARE APPLICATION RECOMMENDATION SYSTEM FOR MOBILE DEVICE USERS

Submitted by Gamze Bayram in partial fulfillment of the requirements for the degree of Master of Science in Information Systems, Middle East Technical University by,

Prof. Dr. Nazife Baykal
Director, **Informatics Institute** ____________________

Prof. Dr. Yasemin Yardımcı Çetin
Head of Department, **Information Systems** ____________________

Assoc. Prof. Dr. Altan Koçyiğit
Supervisor, Information Systems, **METU** ____________________

Examining Committee Members:

Assoc. Prof. Dr. Aysu Betin Can ____________________
IS, METU

Assoc. Prof. Dr. Altan Koçyiğit
IS, METU ____________________

Assist. Prof. Dr. Pekin Erhan Eren
IS, METU ____________________

Assoc. Prof. Dr. Pınar Karagöz
CENG, METU ____________________

Assist. Prof. Dr. Çiğdem Turhan ____________________
Software Engineering, Atılım University

Date: January 29, 2015

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Gamze BAYRAM

Signature: _________________

ABSTRACT

A CONTEXT-AWARE APPLICATION RECOMMENDATION SYSTEM FOR MOBILE DEVICE USERS

BAYRAM, Gamze

M.S., Department of Information Systems

Supervisor: Assoc. Prof. Dr. Altan KOÇYİĞİT

January 2015, 49 pages

Development of smartphones and applications has opened a whole new world for mobile device users. Although this new world has many benefits due to a large diversity, regarding specific application domains, it is getting more complex day by day. In this study, a context-aware application recommendation system that recognizes the situation of users, predicts, and recommends the interactions that are likely to happen by the users in their specific context is developed. The proposed system is based on a hybridization of the Case-Based Reasoning and a Rule Based Reasoning approach that is derived from traditional association rule mining algorithms. Evaluation of the proposed model is done by using a real life dataset collected from individual records of four subjects. These four people were kept track for varying durations from approximately eight months to fourteen months. Results are encouraging when compared to that of previous studies in this domain. Therefore combining these two approaches provides an effective solution to the domain of recommendation systems. To the best of our knowledge, this kind of a hybrid approach has not been utilized in this domain.

Keywords: Recommender Systems, Case-Based Reasoning, Rule-Based Reasoning, Application Prediction, Mobile User, Behavior Patterns, Context Awareness, Personalized Recommendation, Rule Induction

ÖZ

MOBİL CİHAZ KULLANICILARI İÇİN BAĞLAM FARKINDALIKLI UYGULAMA ÖNERİ SİSTEMİ

BAYRAM, Gamze
Yüksek Lisans, Bilişim Sistemleri
Tez Yöneticisi: Doç. Dr. Altan KOÇYİĞİT

Ocak 2015, 49 sayfa

Akıllı telefonlar ve onlar için oluşturulan uygulamaların gelişmesi, mobil cihaz kullanıcılarına yeni bir dünyanın kapılarını araladı. Bu durum zengin çeşitlilik nedeni ile birçok faydayı beraberinde getirmiş olsa da, özel uygulama alanında gün geçtikçe daha da karmaşıklaşmaktadır. Bu çalışmada, kullanıcının içinde bulunduğu durumu tanıyıp, kullanıcının bu belirli durumda mobil cihazıyla yapması muhtemel olan etkileşimlerini tahmin edip öneren bir bağlam farkındalıklı öneri sistemi geliştirilmektedir. Önerilen sistem durum tabanlı anlamlandırma ve geleneksel bağlantı kuralı madenciliğinden üretilmiş bir Kural Tabanlı anlamlandırma yaklaşımın melezlenmesine dayanmaktadır. Önerilen modelin, ölçüm ve değerlendirilmesi dört katılımcıdan toplanan verileri içeren gerçek veri ile denenmiştir. Bu dört katılımcının veri kayıtları yaklaşık olarak sekiz ila on dört ay arasında değişen bir süreç boyunca tutulmuştur. Bu alandaki diğer çalışmalarla kıyaslandığında, elde edilen sonuçlar umut vericidir. Bu yüzden, bu iki yaklaşımı birleştirmek mobil kullanıcı öneri sistemleri alanına etkili bir çözüm sağlamıştır. Bildiğimiz kadarıyla, bu alanda bu tür bir melezleme yaklaşım kullanılmamıştır.

Anahtar Kelimeler: Öneri Sistemleri, Durum Tabanlı Anlamlandırma, Kural Tabanlı Anlamlandırma, Uygulama Tahmini, Mobil Kullanıcı, Davranış Örgüsü, Durum-Farkındalık, Kişiselleştirilmiş Öneri, Kural Çıkarımı

To my beloved family and

To who make me feel loved…

ACKNOWLEDGMENTS

I would like to thank my supervisor Assoc. Prof. Dr. Altan KOÇYİĞİT for his encouragements, guidance and kind understanding during this study. He guided me through all the steps of this study. I am grateful to such an understanding supervisor.

I want to thank to my husband Erman ŞAKACI for being next to me with his support during this period and making life meaningful with his unconditional love.

I would like to thank my mother Meryem BAYRAM, father Alaattin BAYRAM and sister Gizem BAYRAM for their full support during all of the tough journeys throughout my life.

I would also like to express my thanks to the thesis committee, Assoc. Prof. Dr. Aysu BETİN CAN, Assist. Prof. Dr. Pekin Erhan EREN, Assoc. Prof. Dr. Pınar KARAGÖZ, and Assist. Prof. Dr. Çiğdem TURHAN for their valuable suggestions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CBR	Case-Based Reasoning
RBR	Rule-Based Reasoning
HHMA	Hybrid Habit Mining Approach
2D-HHMA	2-Context Dimensional Hybrid Habit Mining Approach
P-HHMA	Partial Hybrid Habit Mining Approach
C-HHMA	Complete Hybrid Habit Mining Approach
AC	Alternative Current
App(s)	Application(s)
GCMP	Generating Candidate promising contexts for behavior Pattern Mining
Wi-Fi	Wireless interfacing for interlinking
SSID	Service Set Identifier
DCON	Digital.Me Context Ontology

CHAPTER 1

INTRODUCTION

The first chapter of this study provides the evaluation of mobile phone technology in a brief manner. After presenting evidence on how significant the mobile technology is a part of our lives, problem in this domain and significance of the study in that manner are stated. It also includes objectives of the study. The last part of this chapter summarizes the contents of the following chapters.

1.1. Background

From a Brick to Many Tricks

It has now been nearly two years since the mobile phone industry has celebrated the 40^{th} birthday of the first mobile phone call in history. The device used on the 3^{rd} of April 1973 by a Motorola employees to make the first globally-accepted mobile phone call was a Motorola DynaTAC, which is known as a "brick" due to its size and weight (Worstal T., 2013).

It was 10 years later when Motorola made this phone accessible for only the propertied class at the time. According to the data from Motorola's website this so called "Dynamic Adaptive Total Area Coverage" system was available for consumers in 1984, weighed nearly 800 grams. Barrett (2010) wrote that this mobile device at the size of brick was sold for $3900 at the time. During the last 40 or so years from the launch of the Motorola DynaTAC, mobile phones have transformed from being a very expensive and hard to reach device to the first thing that we pick-up before we leave our houses.

Figure 1 shows that from 1996 and early 2000 there has not been much increase in the market for mobile phones when compared with the growth in population.

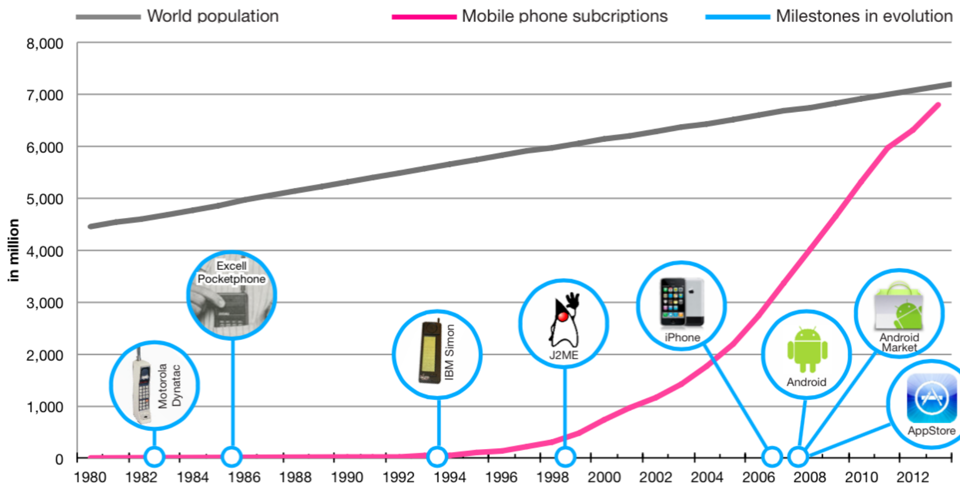


Figure 1: (Matthias Bohmer 2013) Growth of the mobile ecosystem in terms of mobile phone subscriptions in relation to world population.

On the other hand as it is clearly visible from Figure 1 that beginning from the launch of Apple's iPhone devices and the developments of Android straight after, the mobile market has grown significantly to a new direction: Smartphones.

What makes a smartphone "smart" is its ability to run an operating system that is connected to the Internet, which allows the user of the device to download a vast range of applications (apps) to perform multitasking functions similar to what is achievable on a regular computer. Mobile platforms includes Android, iOS, Microsoft Mobile, Palm, etc. Each of these platforms has an application store that bundles all downloadable applications in one area. What these application stores have in common are hundreds of thousands of applications that users can interact with in their daily lives. Figure 2 and Figure 3 show the increase in both present and expected user demand for mobile devices and their operating systems.

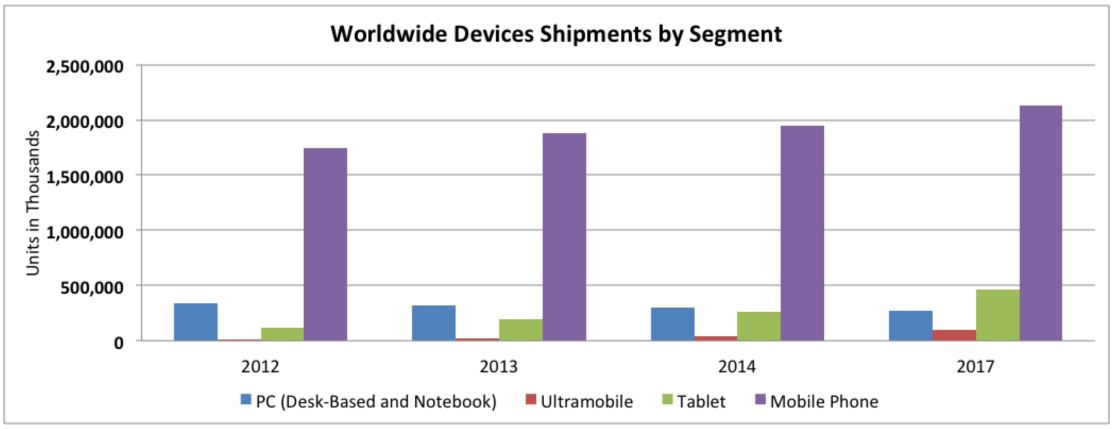


Figure 2: (Gartner 2013) Worldwide Device Shipments by Segment

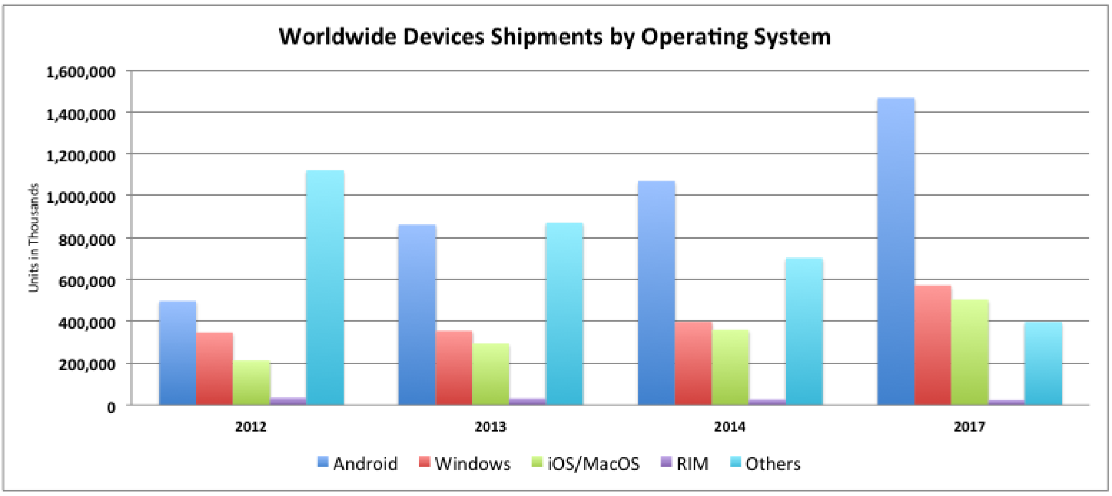


Figure 3: (Gartner 2013) Worldwide Devices Shipments by Operating Systems.

With the introduction of Smartphones and Application Stores, users were introduced with new ways to interact with their mobile devices. Some of these interactions include and certainly not limited to playing games, listening to music, watching videos, reading books, travel planning, taking pictures and utilizing the mobile phone as a GPS device for navigation.

The total usage of mobile devices nowadays is astronomic. Hintze et al. (2014) have found that users interact with their devices an average 57 times resulting in a total session of 112 minutes a day.

1.2. Problem Statement and Definition

Though smartphones are becoming more and more popular, doing even the simplest tasks is getting a lot more complex. This may mean that the user would have to go through a cluster of menu items to get to the required application/ function. In the old days, when somebody pulled a mobile phone out of his or her pocket, the result action would be either a phone call or an SMS (Short Message Service) interaction via the keypad that is always available to access. However, for most modern smartphones with touch screens multiple steps have to be taken to make a call. Steps required to be completed in current vs previous mobile phone designs are compared in the Figure 4 and Figure 5.

Figure 4: Steps required to make a call on current mobile phones

Figure 5: Steps required to make a call on traditional mobile phones

Most of the above mentioned mobile operating platforms may have some sort of shortcuts helping the user to access the desired application, but a phone call is just one of many interactions. One may argue that not all applications are needed all the time, but there are times when we need to reach some applications relatively quickly than any other time. For example, it may not be so important to reach the weather app very quickly but especially when time is very critical, it may be very important to reach the ferryboat schedule straight away at the wharf for our desired destination. A few seconds can make the difference of boarding the ferryboat or having to wait for the next one.

Since it does not make sense utilizing a shortcut for all of your interactions on one screen, there must be another way to make the user interaction with the smartphone a lot "smarter".

An ideal smartphone would know which application the user would require at the time of interaction via recommending these in a list style to make the users experience more efficient.
An important feature of user interactions is that they are generally influenced by the change of the context information. For instance, being in an office during working hours for a workday leads to a different set of interactions than that of in the same time frame on a holiday. This context and user interaction relation has been referred to as *behavior patterns* as per Cao et al. (2010).

Mining behavior patterns allows researchers and developers to understand under what conditions the user interacts with his/her device. Thus leads to the development of applications such as dynamic user interfaces, specialized reminders, multi-functional recommender systems, and even very smart and complex personal assistants are useful in many aspects.

1.3. Objectives

The aim of this study is to propose a system that predicts and recommends applications by thinking in the way the user thinks, thereby saving time the user loses while trying to reach an application on a mobile device and making this experience as simple as possible for the user. Such a recommender system should make life easier rather than making it more frustrated. Especially if this system will follow each footstep of the user it shall also care about how it consumes available resources.

In addition to implementing a system that uses an efficient rule induction approach in order to make it outstanding, our proposed approach compensates the drawbacks of rule induction by utilizing Case-Based Reasoning which is very useful in keeping even specialized and infrequent data.

To the best of our knowledge, in literature regarding the domain of mobile user habit mining there is no study that utilizes such a hybrid approach.

1.4. Outline

This study is organized in six chapters and each of them has the following content.

Chapter 2 includes a literature review on the domain of the study. After stating the approach used in recommendation/prediction systems up to now, mining behaviour which is one of the core elements is explained. Studies proposed on mining behaviour patterns of mobile users and techniques used in these studies are overviewed.

Chapter 3 aims to explain why a hybrid technique is being used. Details about proposed model in both client and server sides and also details of CBR and Rule Derivation is explained in this chapter.

Chapter 4 describes the architecture of the prototype, user interfaces of the related system and working principles of these interfaces.

Chapter 5 starts with the description of the dataset that is utilized for this study. Pre-processing steps applied in order to make data appropriate for the usage of our approach are explained in detail. After that, the parameters used and the train and test phases are clarified together with the details of subjects used in testing phase. According to some performance metrics, evaluation of the study is performed and results are discussed in this chapter.

Chapter 6 concludes the study. It includes discussion the test results obtained from the algorithms. It then finally includes a summary of the core contributions of this work and directions for future studies.

CHAPTER 2

LITERATURE REVIEW

In this Chapter, a literature review on the domain of the study is introduced. After stating the approach used in recommendation/prediction systems up to now, mining behaviour which is one of the core elements has been mentioned. Studies proposed on mining behaviour patterns of mobile users and techniques used in these studies are overviewed.

2.1. Recommendation/Prediction Systems

Some systems adapt to the users' behaviour via examining their behaviours during an interaction in an effort to establish a personalized profile that will be later retrieved in order to select the similar items.

Up to now, different approaches are proposed to achieve such goals. These approaches include collaborative filtering (Chen, 2005), content-based filtering (Zeng, Xing, Zhou, 2003), demographic approach (Krulwich, B. 1997), knowledge based approach (Mandl et al. 2011), and hybrid approach (Porcel et al. 2012).

2.1.1. Mining Behaviour Patterns

Recommendation process starts by learning user preferences. Detecting and modelling interests of the user are primary steps in a recommendation system to achieve personalization in a sufficient manner.

Mining behaviour patterns is essential in order to achieve the implementations of systems capable of making predictions based on previously examined behavioural patterns of the user and thus improving usability of systems via making strategically structured decisions through event detections.

In the literature, it is possible to come across a wide range of domains that mining behaviour patterns has been applied to. In addition to studies in healthcare (Nicolas, Herengt & Albuisson, 2004), education (Ishio et al., 2008), web usage mining (Cooley, Mobasher, & Srivastava, 1999), bioinformatics (Icev, 2003) domains, studies in this area have been proposed in mobile domain too.

2.1.2. Mining Behaviour Patterns on Mobile Domain

As mentioned in the previous chapter, mobile phones are continuing to occupy a growing place in our daily lives. These smart devices that have practically endless capabilities are in our pocket or at an arm's reach distance all day long. Thus, throughout the course of the day people interact in different ways with their devices. For instance, while some people are used to having the news feed open during the times that passes in public transport, most observable habit of some others is listening to music at nights or logging into Facebook in evenings at home to catch up with family and friends. More or less, most people are very familiar with this type of situations. Surely, weekday and weekend usage also makes a difference for our preferred applications on our smart phones. Intuitively, it can be understood that some of these human interactions with smartphones are influenced on the context of the user for certain interactions, as stated by Li et al. (2012). This is a noticeable feature of the human and smartphone interactions. For a system to understand context and act differently according to different context, it must be context-aware.

Context-aware systems can be described as systems that adapt to certain contexts and changes in the state over time. There have been many attempts to define context, many of which resulted in to be too specific according to its domain (Abowd et al. 1999). However, to express context awareness, this study will base its studies in line with the following definition:

"A system is context-aware if it uses context to provide relevant information and/or services to the user, where relevancy depends on the user's task." (Dey, 2001)

Mobile domain is on the cutting edge of pattern mining and makes use of rich source of studies including mobile user movement (Akoush & Sameh, 2007), customer behavior prediction (Eagle & Pentland, 2006), predicting future location of a mobile user (Vu, Ryu & Park, 2009) and mobile commerce (Tarasewich, 2003).

Although there were prediction and recommendation systems available in mobile user domain, only recently some studies that make use of context-rich information and mining mobile user habits in many aspects have emerged. Predictions and recommendations of these studies gain favor in not only a single domain but also in all domains that is convenient to host mobile applications. This is possible by mining mobile application usage habits of the user.

2.1.3 Mining mobile application usage habits

In the literature, mobile context-aware systems that recommend and/or predict applications have been found to be focused on two main goals. The first one is recommending applications to the user that are not currently installed on the mobile device of the related user. The other one is recommending the applications that are already installed and may be preferred by the users, when considering his/her usage habits. In order to achieve these goals, studies utilized from different approaches have been listed in the following lines.

a) Collaborative Approaches

One of the systems that utilized collaborative techniques is AppAware proposed by Girardello & Michahelles (2010). Their system recommends applications that can be installed according to previous selections of other users in the same location for the same type of device in a collaborative manner. Their approach also considers the uninstallation and update ratios of these applications.

Another system having the same purpose is AppJoy. It is implemented by Yan and Chen (2011). Their application returns results based on recently and most frequently used applications by other users in the same context. AppJoy also considers the durations in which these applications are used in that context.

In a study by Bellotti et al. (2008) a system called Magitti has been developed. The proposed system aims to predict what the user will do next on his/her time of leisure and uses context dimensions containing location, time of day, and weather. According to their collaborative study, the authors have discovered other users' preferences is not a good reference in such a personalized domain and decided that for more personalized results users must be able to control what type of context is going to be used by the prediction system and also explored the significance of covering only user controlled context dimensions.

Zhu et al. (2012), mine the meaningful contextual logs from many users and represent each user's contextual features by a distribution of common context-aware preferences. In these logs context dimensions that are given importance are the name of the day, time range (e.g. 8:00am-9:00am), current profile of the mobile device (silent, general, etc.), battery level and location.

Even some research available regarding mobile recommender and predictor systems rely heavily on user ratings, friend based recommendations or weak implicit data such as application installations, according to Karatzoglou et al. (2014), users pay more attention to personalized results rather than results based on recommendations from other users. This case is also supported by Jannach & Hegelich (2009), in their study they have found that personalized results for recommender systems results with an increased number of views and sales, thus attracts users attention at higher levels.

b) Association Rule Mining

Other than collaborative approaches, the literature in this domain has approaches that achieve more personalized results such as association rule mining. Baralis et al. (2011) have developed a framework called CAS-Mine, which discovers meaningful relationships between the users' context logs containing time, date, location, and current interactions with services. Their framework detects generalized association rules in an efficient manner that provides a high level filtration for both user behaviors and services provided. They extract the rules that have been classified into groups in guide with their semantic meanings and this is ranked via quality indices. The system works bilateral for both users and service providers. In this study, researchers evaluated the relationship between some technical thresholds and the number of mined rules.

Cao et al. (2010) and Li et al. (2012) propose an approach that they call *GCPM*. In this approach, they use optimized association rule mining in the domain of behavioral pattern mining. This system captures patterns from users' contextual logs by considering frequency of appearance of applications to calculate the support of a certain context. Based on their experimental results, the authors state that their approach surpasses the traditional rule mining approaches in the domain of mining user behavior patterns. The experimental results were carried out with volunteers who participated for validation. The participants have decided the outcome of the predictions as "yes", "no" or "interesting". In total, 20 interactions were recorded and they have achieved a 95% success rate.

c) **ONTOLOGY-BASED**

Attard et al. (2013) have developed a study in order to enable automatic recognition of recurring events by naming a situation such as "Working@Office", "Footbal" etc. In their study, context information was processed in 3 steps. Filtered context information derived from raw (unprocessed) context data is used to derive situations that are recurring sets of unprocessed and filtered context patterns. In their approach, an ontology-based graph matching technique was utilized. Their ontology, which they call DCON, is one of the most comprehensive ontologies available in mobile context domain.

d) **NAÏVE BAYES**

Kamisaka et al. (2009) proposed an approach which is composed of Naïve Bayes classifiers and zero-attribute role (ZeroR) in order to examine application usage of mobile users. In their study they have defined Naïve Bayes (NB) as the following:

"NB is a simple probabilistic classifier based on Bayes' theorem with strong (naïve) assumption that the attribute values are conditionally independent given the target value." (Kamisaka et al., 2009)

ZeroR however, is a less complex most frequently used based method. In their study they have compared these two models to determine which method is superior to the other. Their logged data consists of the *actions* that are executed by the user, *events* such as incoming calls and alarms, *location* received by a GPS chip, *flip status* for clamshell formed mobile phones that indicate if the screen is open or not, *signal strength*, *battery power level* and *silent mode status* which indicated if the phone is on silent mode or not. The authors have filtered out the weekends from the records in order to keep only the weekdays assuming that people behave in a more periodical way on weekdays rather than other days of the week. On top of this they have also excluded the contextual data of the user when the user was away from his/her office by 50km, considering that this situation is unusual and the user is assumed to be on holidays. This is a poor assumption when for example the user is a sales representative of a company that travels to different suburbs to visit his/her clients or a medical practitioner that works on a rotational shift, which means that there is neither weekday nor weekend for this type of user. The study involved 19 subjects to be followed for a period of 1 to 8 months. Their approach was evaluated with 9 applications amongst 15 candidate apps. Their evaluations are mainly based on detecting which context dimension based recommendation was more effective on the decision making process. They also distinguished interactions that could be learnt by their approach in training phase from the ones that could not be learnt.

In another study completed by Lee, Choi and Kim (2011), an Android application recommends a number of applications that best matches the user's current context that consists of time, location, weather, activities and emotion. The authors use a supervised machine-learning approach utilized from Naïve Bayes classifier and creates a probability model. Their study involved the participation of 2 subjects for a period of 2 days. During this time they have retrieved 163 cases in total. The authors have achieved 69% accuracy to determine 1 application out of 3 application candidates.

Bridle and McCreath (2006) also have studied Naïve Bayes based approach in encouraging a user to select shortcuts for phone and text messaging in order to save the total number of buttons used to reach these applications. For their approach they have used real-life data recorded on Nokia's Symbian platform. In their study they have evaluated four methods and a hybrid model that is based on two of these methods. The methods evaluated are *Naïve Bayes* based, *Decision Tree Based*, *Fisher's exact test* based, *most frequently Used* based and the hybrid method utilizing *Naïve Bayes* and *Most Frequently Used*. However in their study they have only targeted SMS and Voice-Calling applications that is only a very small fraction of what people use on their mobile devices. This insufficient set of applications is not enough to realize context-aware user interactions. Predicting the most commonly used applications is a good way to start prediction for a given user, however predicting the infrequently used applications for a given context is more crucial and a bigger challenge in mobile application prediction systems.

In the study by Shin, Hong and Dey (2012), they argue about the fact that when the number of applications is increased in a mobile device, it gets harder for the user to find the location of the shortcut of any desired application. For this reason, they have aimed to develop the most accurate prediction system for the smallest number of applications to be shown to the user. To achieve this they have created an application based on personalized naïve Bayes methods for each user to predict applications. They use GPS, time, Battery, Last used applications, Cellular Networking, Settings (Ringer volume, system volume, vibrate mode etc.), 3D accelerometer, illumination, screen status, call-SMS events, Wi-Fi and Bluetooth connectivity status as part of their contextual data. In order to use probabilistic approaches they discretized their context in five main statuses such as: very low, low, medium, large and very large. The authors have tested their approach on 23 users for a month. At the end of their studies the authors have evaluated that their model has a correct prediction rate of 65% - 75% for 5 candidate applications out of 32 applications.

Kurihara, Moriyama, & Numao (2013) have pointed out a common issue with recommendation systems. They have given a typical scenario for a user who wants to review current information regarding train schedules at a train station. In another scenario they mention about how users only check the news and/or weather application in the morning and never return back to it. These typical scenarios suggest how some applications are not frequently used even they are more context-dependent. Researchers indicate that users would feel rather stressed to try to find these applications that they do not use in short periods of time. The authors state although these applications are not used frequently they are still worth recommending at certain contexts. To predict behavioral patterns the authors have mentioned that statistical methods can be used however, these methods cannot understand infrequent patterns such as the scenarios that they initially presented. Their aim is to tackle this problem via an approach they have called "*event frequency – inverse context frequency*" that is used to define applications that depend on context. Since the frequency of certain words (keywords) follow a power law, they argue that their method extracts these methods very effectively.
The evaluation of their approach involved testing 10% of their data against 90% training of the data they have collected from five individuals. Their calculated success rate were 53.8%, 72.9%, 50.2%, 66%, 65.4% respectively.

2.1.4. Case-Based Reasoning in Mining Behaviour Patterns

The term "Case Base Reasoning" can be traced back to the study by Schilit and Theimer (1994) where this term had been used for the first time. The authors have defined CBR as a systematic approach to extract relevant cases according to the users' current situation. Knowledge in cases covers issues regarding the relations between certain items and certain user needs. This way they can make logical assumptions about the relationship between needs and recommendations (Bruke, 2002).

Since then there has been extensive research on Case-Based Reasoning (CBR) including some that are in the mobile device domain. The majority of these studies are based on domain specific applications such as a study by Lee and Lee (2007). It focuses on presenting users an application called "C2_Music" that was developed using real world data to recommend a certain type of music for a specific user given in a specific context. This application looks back into the habits of similar users to recommend the same type of music.

Problems with CBR

Disputes about CBR gather around the fact that without statistically proven data for supporting its retrievals, the correctness of the generalization cannot be guaranteed. It is necessary to formalize case-based inference in order to allow it to generate case-based predictions supported with a certain level of confidence.

Another criticism on CBR is about the complexity. Case Based reasoning is a connectional method used for problem solving and uses past knowledge to solve present problems. Traditional CBR methods have to evaluate all the cases stored in the case-base in order to return the most relevant case(s). For this reason the efficiency of the CBR method is negatively affected due to the size of the case-base.

Compensation techniques used in case-based approach

In literature, there are studies in which hybrid algorithms are used in order to add inference functionality to CBR. One of these studies is the study of Schiaffino and Amandi (2000). They have initially described the following to better understand the concept. In their study, they consider a user who is a student at a university that utilizes a database across its information resources. This user sends queries to the database and interacts with the system via different tasks to get the information he/she is after. This can be information about careers, departments, other students, coursers or marks. Their aim is to develop an algorithm that determines exactly which knowledge the user requires and helps users to reach information that is most relevant to them. Their paper present a method that integrates above mentioned Case-Based Reasoning and Bayesian Networks in order to build personalized user profiles. CBR provides the infrastructure to keep the knowledge about user interactions that are meaningful in order to determine a user's habits and preferences. On the other hand, Bayesian Networks is useful to build a relation between quality and quantity and items of interest. The authors have stated that they source the required knowledge set to build the Bayesian Network from the stored cases in the case-based reasoning section of the algorithm. In their technique, they use CBR as a "slave". This is due to fact that cases recorded in the CB section of the algorithm is used to calculate the probabilistic values associated on each note of the Bayesian Network.

In a study by Liu, Ke and Wu (2008) the authors adopted rule inference and information retrieval techniques to a case-based reasoning approach designed for the purpose for situation recognition on a production line. In an effort to understand the frequent associations in between meaningful context features and situational features, association rule mining method is used to discover context-based inference rules. Combining two features in order to develop a system is referred to as hybrid based systems.

Compensation for the complexity problem in case based reasoning, even some indexing mechanisms to make a CBR work efficiently have been proposed, recently a new and more promising approach is suggested. Beaver et al. (2013) proposed a system that derives and uses knowledge about preferences of the user in run time during a conversation on selecting a flight. This system utilizes case-based reasoning to obtain preferences. MongoDB's flexibility in storing data and ability to scale even if huge numbers of cases are recorded are stated as the reasons in using this database. MapReduce mechanism that is very useful in complex and parallel searching among cases was one of the principle reasons in preferring such a database for the authors. They stated that the total average user time required to conduct a search in the database met their expectation on serving users in real time.

CHAPTER 3

PROPOSED MODEL

In this chapter after recalling alternative models in the literature, the reason to choose such a hybrid model is stated. Details about proposed model in both client and server sides and also details of CBR and Rule Derivation are explained.

3.1. Alternative Models in Literature

Recommender predictor systems proposed up to now are developed with different approaches. One of these is content-based recommender systems, which recommends items based on the content that the user may be interested in. Although this approach brings in benefits such as user independence and recommending items even if any user has not yet rated them, it has important drawbacks on the domain that we are studying. First of all, if the content does not contain sufficient amount of information, system that utilizes this approach cannot select the items that user likes among the others that user does not like.

Another alternative is *Collaborative Recommender Systems*, which utilizes common interests rather than personalized ones. The recommendations of these systems are based on reviews of users who share similar interests with the user who is receiving the predictions. There are some drawbacks commonly associated with collaborative recommender systems. One of these is that the learning curve of collaborative recommenders is too steep. It takes a long time for collaborative systems to understand what the user preferences and ratings are. Collaborative systems also deal with inadequacy problems. This is a problem faced when very few people rate a lot of items in the ecosystem. Even though a lot of items are rated, since these have only limited reviews, collaborative recommendations systems tend not to recommend these items. Unusual users may also be a problem of ecosystems where there is only a small or even medium community of users. Individuals may receive recommendations that do not reflect their choices since their opinions do not match with any other people in the ecosystem.

There are also *Demographic Recommender Systems* available in literature. Such systems' main goal is to categorize the user as a profile according to the users personal attributes thus allowing recommendations to be made according to demographic classes. These methods are known to have less success rates than the rest (Gemmis et al., 2011).

Another study in this area is referred to as *Knowledge-Based Recommender Systems*. These systems rely on knowledge-based approaches to recommend items to users. Knowledge- based recommender systems do not suffer from problems such as the early rate or “too many” rate problems that is associated with collaborative systems. It can be stated that all systems that generate recommendation as a result can be referred to make some kind of inference. CBR is a technique utilized in such type of recommendation systems.

3.2. The Reason To Choose Case Based Reasoning and Rule Based Reasoning

Since habit mining is a user centered process and has to behave differently for different users, the method used to recommend applications needs to adapt to different cases. Due to the fact that case-based reasoning is capable of extracting specialized knowledge it makes sense to use CBR in this domain. One of the unique capabilities of CBR is that it can update itself according to the new cases it faces during real-life situations. It can do this since each case is stored and removed from the knowledge as individual modules. These features allow CBR to be highly maintainable and easy to adapt to different situations. However the specialized nature of CBR can be a disadvantage at some times since it is not capable of drawing general knowledge. Moreover, when the case-based grows over time, it takes more time for the CBR method to find and match the best cases according to context. On the other hand, rules are good at generalizing human habits and are highly efficient method to be utilized on mobile devices.

If the advantages of CBR and RBR are utilized and trade-offs are eliminated at a certain degree, the combination of the two methods can offer significant benefits. The combination of these two methods is also justified by the fact that these two methods mimic the complex nature of human thinking. When facing a new situation an experienced person combines general knowledge (rule based) with experience-based knowledge (case-base) in order to solve a problem. Combining two methods is referred as a *hybrid system.* The aim of this approach is to develop a system composed of two different sub-systems that take advantage from each of their components. In general terms, it is accepted that solving larger and relatively complicated problems is easier with a combined approach. Prentzas and Ioannis (2007) have stated that the utilization of Rule Base Reasoning and Case Based Reasoning is an effective solution to solve even the most complex problems:

> *“The effectiveness for these approaches stems from the fact that rules and cases are complementary in representing application domains and solving problems”* (Pretzas and Ioanis, 2007)

3.3. Details about Proposed Model

CBR considers reasoning as a technique to remember one or a limited set of discrete events or cases and bases its decision making process according to comparisons between newly faced situations and stored ones (Schiaffino & Amandi, 2000).

3.3.1 Case Structure

Every interaction logged by the user is represented in a case form. For this reason a case is referred to a discrete piece of knowledge that contains an experience. All recorded cases in the case database teach the system a session, which forms the fundamentals of achieving the goals of the algorithm that will use it. In this study, a case has two main parts. First part describes the situation or the problem and the second part returns a solution for the described problem.

Figure 6 presents the structure of a case. As per the figure, a problem covers the user, timestamp and context. In the context domain, “Day of the week” is defined as weekdays or weekends. For the “Period of Day” a day is split into four groups in which the details can be seen in Table 1.

```
Case

            Problem
                User:
                      {User Name}
                Time Stamp:
                      {yyyy-mmdd-hhmm (eg., 201501291300)}
                Day of Week:
                      {Weekend, Weekday}
                Period of Day:
                      {Morning, Afternoon, Evening, Night}
                Battery Level:
                      {1, 2, 3, 4, 5}
                Charger:
                      {AC, USB, Disconnected}
                Activity:
      Context             {Stationary, Walking, In Vehicle}
                Connection:
                      {Mobile, Wi-Fi, Disconnected}
                Wi-Fi (if Connected):
                      {SSID}
                Location:
                      {Home, Work, Out}
                Airplane Mode:
                      {On, Off}
                Ringing Mode:
                      {Vibrate, Silent, Normal}
            Solutions
                Interaction:
                      {Name of Interaction}
                      {Detection Count}
                      {Last Seen}
                .
                .
Interaction List .

                Interaction:
                      {Name of Interaction}
                      {Detection Count}
                      {Last Seen}
```

Figure 6: Case structure upon a query

Table 1: Time Periods of the Day

Morning	Afternoon	Evening	Night
06:30 - 12:29	12:30 - 18:29	18:30 - 00:29	00:30 - 06:29

Battery level is split in a similar way where every number represents a rate of values. The value of each battery level is shown in the Figure 7.

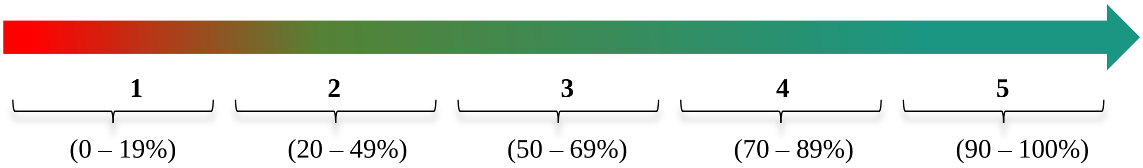


Figure 7: Representation of Battery Levels

Connection means the status of the connection depending on its type. If mobile internet (i.e., EDGE, 3G, 4G, etc.) is active, it is recorded as “Mobile” otherwise the case either contains “Wi-Fi” or “Disconnected” depending on the current situation. When connected via Wi-Fi, case also contains the SSID of the Wi-Fi connection point for future reference. Location however is determined by using GSP coordinates of the user at the current context. If this value was pre-defined as “Home” or “Work”, this definition is recorded in the case; otherwise, the current latitude and longitude recorded as a third class. Airplane mode is a binary class and recorded as “On” or “Off” accordingly. Ringing mode represents preferred ringing status. The case contains the solution list (i.e., app interactions) together with the problem itself. Solutions in case are listed by considering how much they are interacted in the related context. As a second parameter, last seen time is used in case of count of detections are the same.

3.3.2. Workflow in the Proposed Model

The workflow in the proposed model can be grouped into two sections. These sections are *server side* and *client side*.

3.3.2.1 Workflow in Server Side

In the server side, a case-based and a rule-based module are incorporated together. In server side, there are four main cycles that cases passes through. When a case is formed immediately after detecting any change in users’ context or an interaction with any application, this case follows the steps seen in the diagram in Figure 8.

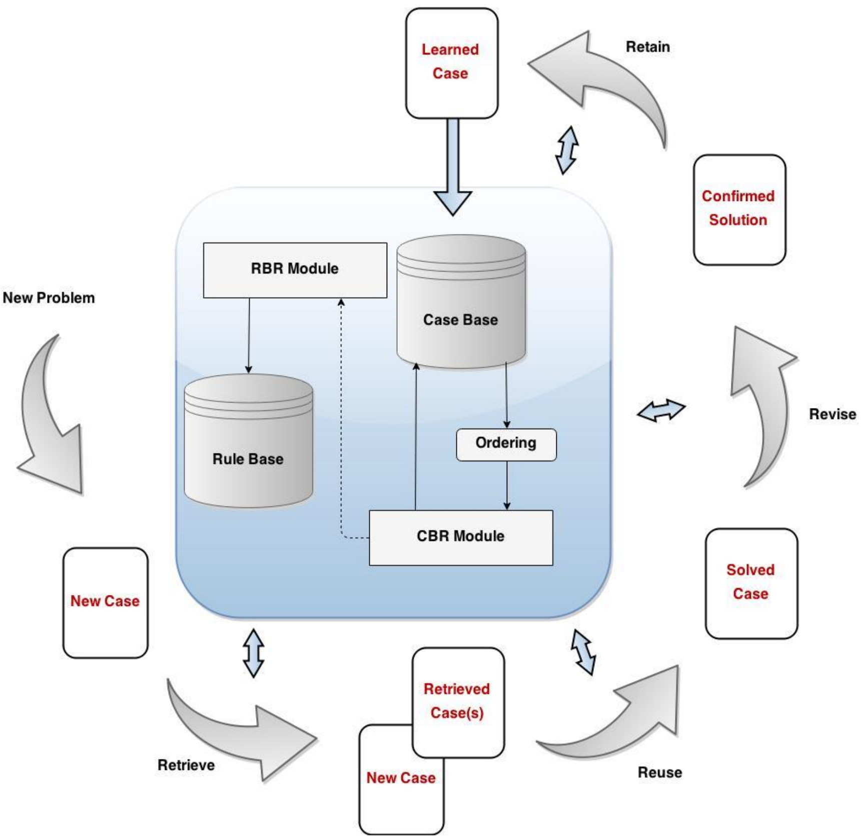


Figure 8: Case-Based Inspired Hybrid Cycle

The cycle seen in Figure 8 represents the 4R's (Retrieve, Reuse, Revise, Retain) that are the fundamentals of case base reasoning. For both the case base reasoning and the rule base reasoning modules, the most similar cases are retrieved. If similar cases are available, these cases are re-organized in order to form a new solution list to be recommended. If the case was formed because of a change in users' context, confirmation of user about the recommendation is received. The confirmed solution is retained and related case is updated. If a user selects an application that is not available in the prediction list, this new selection is retained and the related case is revised. The sub processes are presented in Figure 9.

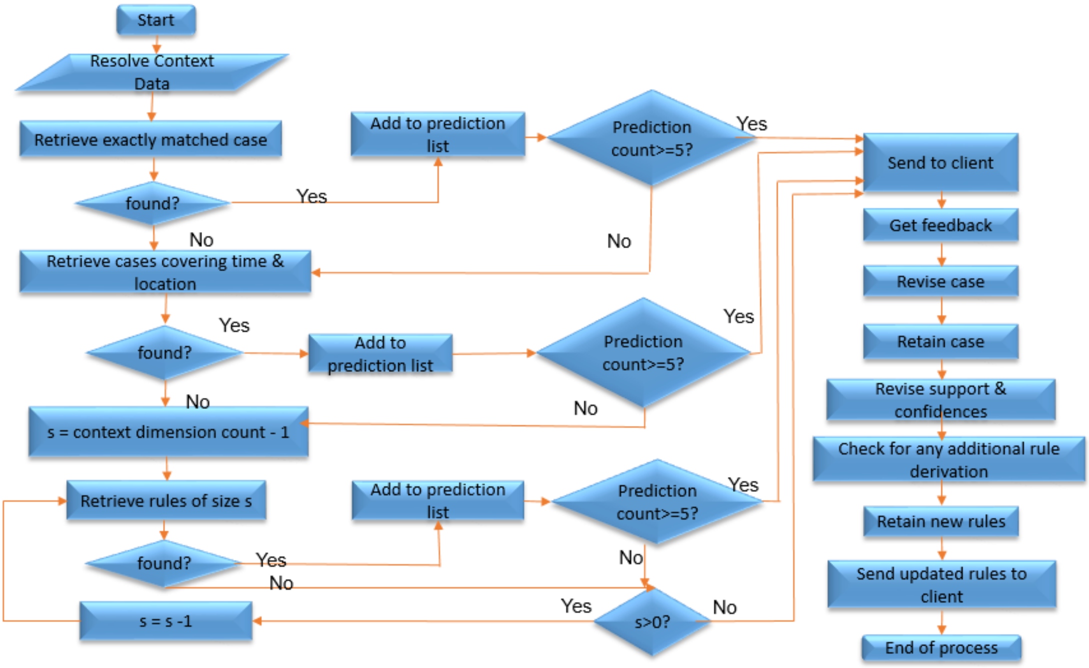


Figure 9: Processes in server side upon a user demand or change in user context

3.3.2.1.1. Retrieving and Reusing Phases

When a new case formed as a result of a change in user context, a user demand or an interaction detection has been sent to the server, this case is split into its features and searched in the case-base. While querying the case base for an existing similar case evaluation of all the context dimensions are considered as equal in the first phase of the retrieval process. In case of finding any exact match with this case, solutions that were listed in this case are added to the predictions list in the same order. The number of interactions per an application has determined this order in a decreasing manner. In case of a situation where equally detected interactions are found, the interaction with the newest timestamp will be recorded to a higher level on the list. If the previously defined number of prediction count is achieved, the prediction list is sent to the client. Otherwise the second phase of retrieval is executed. In this phase, time is considered as an essential dimension as Saleh & Masseglia (2011) have stated. Also as examined in literature review, previous studies in the same domain use location together with time. Thus, it is safe to say that time and location are the most important features of context and it is also safe to use both dimensions in the second phase of the retrieval. Hence, the case base is queried once more. This query will require only the exact match of both time and location dimensions. The case that was recorded in the same time slot, in the same day type (weekend or weekday) and at the same location is retrieved. Solutions that will be used in this phase is ordered by considering their total occurrences, in case of equality, time stamps are considered in the same manner. If previously defined prediction count is achieved, prediction list is sent to the client. Otherwise, rule based search is performed. The technical details of how the rules are inferred will be explained in the following sections.

If the time and location dimensions in the context that are currently being searched have not been previously seen in a similar case, rule-based search will return an empty result, whereas the other context dimensions may be able to give an important clue to be recommended. Inferred rules containing one less context dimension than the context that is being searched is retrieved from rule base if available in any combination in this count of dimensions. Again, previously defined prediction count is checked and in case of no achievement, the rule-based module decreases the count of dimensions of the context to be searched by one. This process continues until either previously defined prediction count is achieved or size of context to be searched becomes zero.

3.3.2.1.2. Revising and Retaining Phases

After a prediction list is sent to the client, the system receives the selection of the client. According to this selection, the case is revised. After revision, if the preferred application does not exist in the solution list of the case, it is added. If exists, its count of detection in this specific context is updated together with frequencies of all dimensions of context an interaction listed in the solution list. Then, the retained case is stored in the case base. Case-based module also triggers the rule-based module and then supports and confidences about the context that is clarified in the following section, are also revised together with the selected interaction. If it is possible to derive any additional rule combination by using revised support and confidences, these additional combinations are stored in the rule base. After sending updated rules to the client, the process is completed.

3.3.2.2. Process in Client Side

The processes that the client side is responsible for covers two main processes. The first one is, if client is connected to the server, it will display the prediction list coming from the server. This list is formed by utilizing the hybrid approach. The second one is valid in case of any connection problem with the server. In this situation, client searches the best solutions to the current problem among the pre-configured rules. The approach that is utilized in this stage is only the Generating Candidate promising contexts for behavior Pattern Mining (GCPM) (Cao. et al. 2010) approach and it provides keeping track of users and assists them even when connection is lost.

In case of a user demand or a change in context of the user, client side process starts. The steps of this process are shown in the following Figure 10.

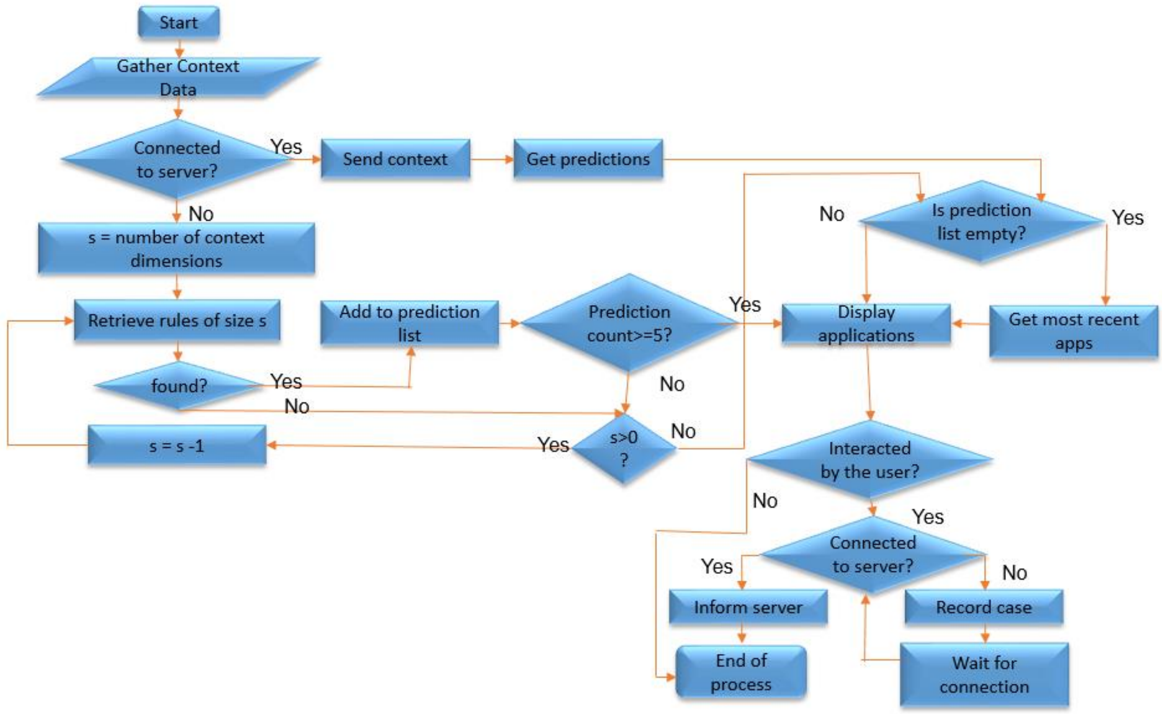


Figure 10: Processes in client side

As it can be seen from Figure 10, gathered context data is sent to the server if connection to the server is established. If prediction list sent from the server is not empty the applications in this list are displayed. If any application is interacted by the user, server is informed about this application and the process ends. In case of an empty prediction list, applications to be displayed will be the most recent applications of the client.

However in case when there is no connection to the server, rule based search is performed among previously recorded rules derived and sent from the server. The first thing to be searched is to find if there is a rule of which size is equal to the size of gathered context which covers the same features and same values. If found, it is added to prediction list. Until previously defined prediction count is achieved or size of the context to be searched becomes zero, decreasing this by one and adding the interactions recorded in the found rules continues. Once prediction list is formed, applications in this list are displayed to the user. If any interaction by the user is detected, and connection to the server is still not achieved, case is recorded in order to inform the server when connection is established again. Otherwise client informs the server and process ends. In the case of empty prediction list, applications to be displayed are determined the same way. These applications are the most recent apps used by the client.

3.3.3. Rule Derivation

Successful cases that are retained for usage in feature problems are used to derive rules. In order to derive rules we were inspired from the study of Cao et al. (2010).

The authors called their approach GCPM, which is an efficient approach to mine behavior patterns and is derived from traditional association rule mining algorithms. The definitions of some of the key points in this approach are given below in parallel to the study Cao et al. (2010).

CONTEXT

Consider a contextual feature set $F = \{f_{1,} f_2, \dots, f_k\}$, a context C_i is defined as the contextual feature-value pairs such as, $C_i = \{(x_1 : v_1), (x_2 : v_2), \dots, (x_l : v_l)\}$. Here $x_n \in F$and v_n as the value for x_n $(1 \leq n \leq l)$. A context with l contextual feature-value pairs is defined as a $l - context$.

SUB-CONTEXT, SUPER-CONTEXT

Consider two contexts C_i and C_j. For this pair if $\forall_{p_i \in C_j} \in C_i$, where p_i refers to the contextual feature-value pair, C_j is called a sub-context of C_i and C_j is called a super-context of C_j.

On the other hand a contextual feature refers to a context data type. This may be the period of the day, location, ringtone state, etc. The contextual feature-value pairs are stored in an order predefined according to contextual features.

INTERACTION RECORD

An item in the interaction set (Γ) is referred to as an interaction record. Here, $\Gamma = \{I_1, I_2, \dots, I_Q\}$, and I_n $(1 \leq n \leq Q)$ means an interaction by the user. Interaction records are the recorded interactions the user encountered with the mobile device such as playing a game, making a call or browsing the web.

CONTEXT RECORD AND CONTEXT LOG

Given a $r =< Tid, C_i, I >$ this context record consists of a timestamp T_{id}, a context C_i and a user interaction I.

A context log however, is a group of these context records in the order according to their time stamps i.e. $context$ log $R = r_1 r_2 \dots r_{N_R}$

SUPPORT AND CONFIDENCE

For a context C_i, an interaction record I, and a context log R, the support of C_i with respect to I (shown as $Sup(C_i \Longrightarrow I)$) is denoted as the following:

$$Sup(C_i \Longrightarrow I) = \sum_m Count_m(I)$$

Here, $Count_m(I)$ is equal to the total times the interaction I has been seen in the m-th context range of C_i.

The support of C_i $(Sup(C_i)) = \sum_{I \in \Gamma} Sup(C_i \Rightarrow I) + N_0$, here N_0 denotes the number of context ranges according to C_i that do not contain any non-empty interaction records.

The confidence of C_i with respect to I ($Conf(C_i \Rightarrow I)$) is shown as $\frac{Sup(C_i \Rightarrow I)}{Sup(C_i)}$

PROMISING CONTEXT AND BEHAVIOR PATTERNS

For a context C_i and two user defined parameters *min_sup* and *min_conf*, if $\exists_I\ Sup(C_i \Rightarrow I) \geq$ *min_sup*, then C_i is called a promising context.
Moreover, if $Conf(C_i \Rightarrow I) \geq$ *min_conf* in this case $Ci \Rightarrow I$ is called a behavior pattern.

GCPM

The main approach of GCPM is to join promising l-contexts.
Considering the following contexts:
$C_i = \{(x_1 : v_1), (x_2 : v_2), \dots, (x_l : v_l)\}$ and $C_j = \{(y_1 : u_1), (y_2 : u_2), \dots, (y_l : u_l)\}$
If $\forall 1 < n < l\ (x_n = y_{n-1} \wedge u_n = \wedge u_{n-1}$ then C_i and C_j are said to be *join.*
The joint context of C_i and C_j, $(C_i \cdot C_j)$ = $\{(x_1 : v_1), (x_2 : v_2), \dots, (x_l : v_l), (y_1 : u_1)\}$.

For a context log R and a user interaction set $\Gamma = \{I_1, I_2, \dots, I_Q\}$, GCPM's first approach is to sort all l-contexts seen in R as $\wedge^1 = \{C_i^1\}$ and set l to be 1.

CHAPTER 4

PROTYPE IMPLEMENTATION

In this chapter architecture of prototype, user interfaces of the related system and working principles of these interfaces are explained in detail.

4.1. Architecture of Prototype

For the proposed prototype 3-Tier architecture, which consists of 3 layers, is illustrated in Figure 11.

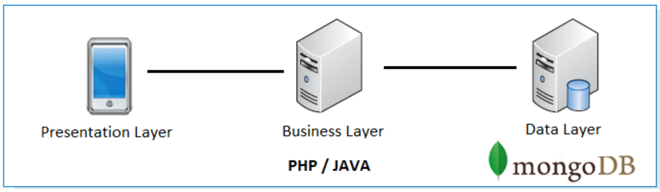


Figure 11: 3-Tier Architecture

The first layer is the Presentation Layer. This layer is responsible for tracking user context and interactions, passing this data to the Business Layer, and displaying the prediction that is derived either by waiting for new knowledge from the server or using previously defined knowledge in itself. In the proposed model, the Presentation Layer is implemented in Android Platform.

The Business Layer handles the processes that form the service given by the server. This layer handles the data coming from the presentation layer in order to derive knowledge from this data and transmits the knowledge to the data layer to store this information. For this layer, PHP and Java is utilized in the development.

The Data Layer is the third layer of the 3-tier architecture. It stores knowledge derived by the Business Layer. This knowledge includes derived cases and rules. It can also re-access this data to read, delete and also update it. For the Data Layer, MongoDB has been used. The reason for using MongoDB is the aim of utilizing NoSQL in order to store objects as JSON objects. Benefits of using MongoDB have been argued by Beaver & Dumoulin (2013). As mentioned by the authors, MongoDB is highly flexible and at the same time easily scalable when the number of cases increases.

Also its indexing, partial and complete search mechanisms as well as mechanism that allows complex parallel searching are big advantages for not only the study of Baever and Dumoulin (2013), but also all cased based approaches such as our approach. Test results of this mentioned study proved that even in very large case bases, near real time response is provided by utilizing this mechanism.

4.2. User Interfaces and Working Principles

In this section, the user interface of the prototype is described.

Figure 12: Welcoming screen

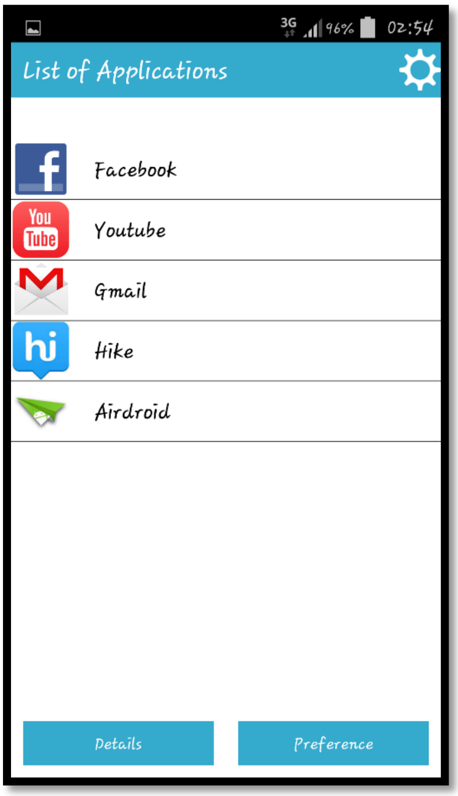


Figure 13: Prediction list

When users launch the application, they are greeted with a welcoming screen according to their contextual information. Application displays a pop-up dialogue box informing the user that it is retrieving the applications list according to the information processed Figure 12. Once the relevant applications are displayed on the user screen, the new page will be similar to the one seen in Figure 13.

At this point the user is at the application prediction/recommendation page of the application. From here, users are able to reach all the actions available on the application. The options available are the “Set your Location” icon at the top right hand side of the screen, the “Details” button at the bottom left hand side of the screen and “Preferences” at the bottom left hand side of the screen.

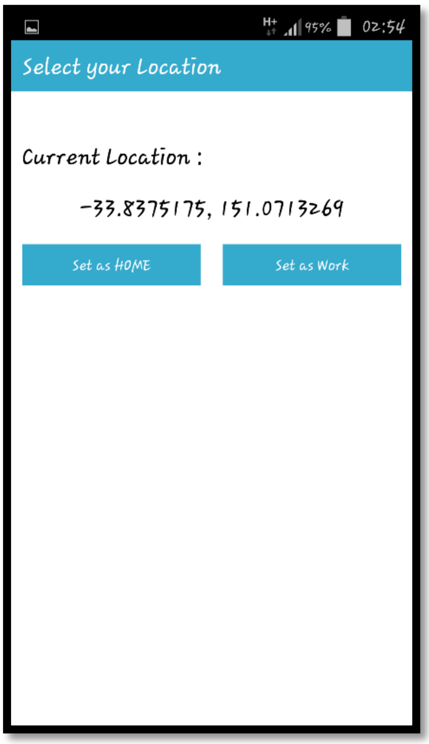


Figure 14: Context before setting home or work

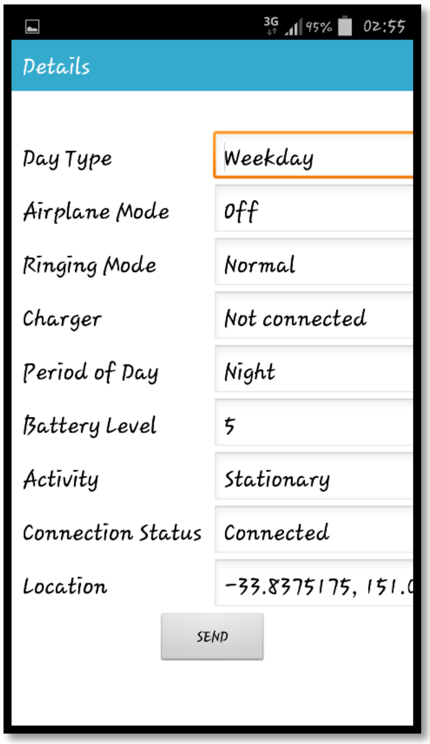


Figure 15: Setting home or work

On the instance when a user selects the “Set your Location” icon, the new screen will be as in Figure 14. Users are allowed to perform three actions. They can either set the shown coordinates as home, work or the user can return to the previous page by hitting the Android “back” button.

After users return to the main menu, and select “*Details*”, this time users will see contextual data gathered by the application as in Figure 15. From this menu, users can alter the information shown and/or can manually “*Send*” their contextual information to the server.

Note: Assuming that users have started the application for the first time or they have not set a location, their “*Location*” information box will show the users’ coordinates or the corresponding value (i.e. Home or Work).

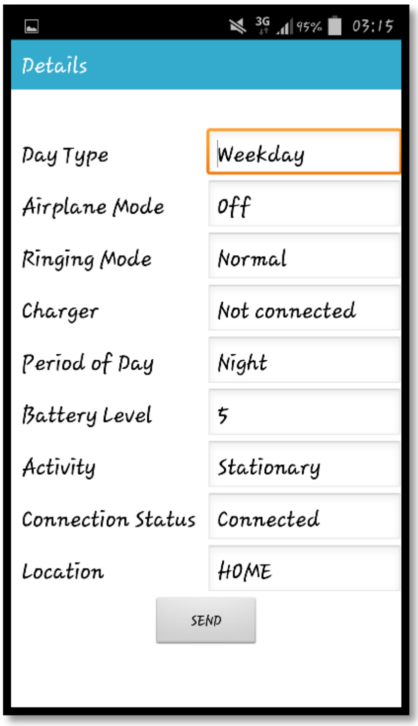


Figure 16: Context after setting home

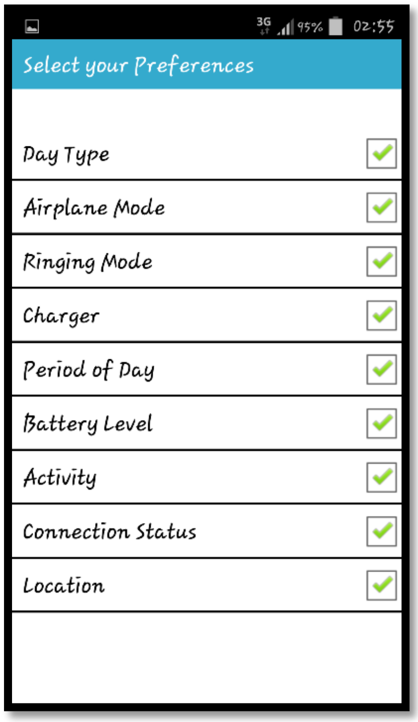


Figure 17: Context to be sent

The last settings that the user has control over is the "*Preferences*" screen. Here, users can select which contextual data that they allow or prefer to be sent to the server to be stored in cases (Figure 17). The importance of such feature has been stated in a study by Belletti et al. (2008). He argues that rather than having a large amount of contextual data, it is more important to use contextual data that makes more sense in individual base. For that reason, allowing the user to select the most appropriate context dimensions is proved to be an important feature.

CHAPTER 5

PERFORMANCE EVALUATION

This chapter starts with the description of data set that is utilized for this study. Pre-processing steps applied in order to make data appropriate for usage of our approach are explained in detail. The parameters used and train and test phases are clarified together with the details of subjects used in testing phase. According to proposed performance metrics, evaluation of the study is performed and results are discussed in this chapter.

5.1. Explanation of Data Set

A good dataset is essential to get results as close to real world situations as possible. There has been some contextual data collection tools developed with different results. For instance Falaki et al. (2010) developed a custom logger to collect information from 33 Android and 222 Windows Mobile users. In their study they have logged information such as the screen state, used app, networking status and battery state.
Another tool developed by the University of Oulu (Finland) is called "AWARE" which is an open source data collection tool developed by the Department of Computer Engineering. It encourages developers to use their tool to collect contextual information in order to code smarter applications that are aware of the users' context. Though this is an intelligent framework that is easy to adopt thanks to the Android library that they share, the university does not share the context collected at their database.

Similar to AWARE, there is another framework called "Funf" which has been developed by a project of Human Dynamics research group at the MIT media lab. Eagle and Pentland (2006) collected data from 100 mobile phones over a period of 9 months.

Another study in this area by Rawassizadeh et al. (2013) is UbiqLog that is a context-scanning tool that was developed with the real-world data of 6 users spanning a time between one to fourteen months.

The last context-collecting tool that will be mentioned in this report is "DeviceAnalyzer" by Wagner, Rice, and Beresford (2013). The study that has been conducted currently contains the biggest set of Data available upon request for all developers and researchers. The database available for development has taken measures to protect against identity detection via encrypting sensitive data and includes the data of 12,500 Android Devices in the wild for a time spanning of nearly over 2 years. This is a very important study where anyone can participate to become a part of a project that aims to be the biggest “real-life” information database available. In this study, records available from this database is used to train our approach and to evaluate how our prediction engine will behave in real world circumstances

5.2. Data Pre-Processing

In user habit mining, capability of dataset in reflecting the reality and cleanliness of data determines its level of accuracy. For this reason, pre-processing of the data carries a big importance.
In order to obtain results as close to real life as possible pre-processing of data is performed by following the steps given below.

5.2.1. Elimination of Less Significant Attributes

The data set of DeviceAnalyzer project is a huge dataset that can be processed in a lot of different domains. For this study data to be utilized is chosen according to the aim of the study. The context dimensions that is used for this study as it is stated in the following table.

Table 2: Derived contextual data from dataset

Sensor	Contextual Data	Possible Values
Time	*time|bootup**	{*yyyy-mm-ddThh:mm:ss.sss+0100*}
	*This is the value of local time when the system booted. This value is used as a reference to calculate the local time of when an application started.	
Setting	airplane	{*on, off*}
	audio|ringermode	{*normal, silent, vibrate*}
App	app|recent|n**	{app_0, app_1, ..., app_n}
	**n represents a value such as 0, 1, 2, …, etc.	
Connection	conn|WIFI|detailedstate	{*connected, disconnected*}
	wifi|scan|{SSID value of WiFI}|ssid	{*SSID value of WiFI*}
	conn|mobile|detailedstate	{*connected, disconnected*}
Image	image|dates***	{*array of values*}
	***image values are a comma-separated array of millisecond timestamps that contain information of when the relevant picture was taken. This value changes whenever the library is modified such as a picture being added, or removed. In this study we have assumed that the user executed a camera interaction with the phone whenever this value changes.	

Cell Location	phone|celllocation|cid (Cell ID)	{*cell ID* or -1 if unknown}
	phone|celllocation|lac (Location Area Code)	{*LAC* or -1 if unknown}
Power	power|battery|level	{*percentage value of level*}
	power|charger	{*usb, ac, disconnected*}

5.2.2. Handling Missing Values

The working principle of DeviceAnalyzer is based on two different data obtaining techniques. Some of the data is recorded whenever a change is observed whereas the others are gathered by scanning the mobile device in some pre-defined intervals. In order to take different and sufficient number of context dimensions into account we assumed that the valid value for each context dimensions are the value that is the last seen against the related dimension.

5.2.3. Data Transformation

In data set of DeviceAnalyser, some of the data is scanned in pre-defined intervals and is recorded together with the time elapsed since boot-up of the device in millisecond format. Transformation of this millisecond value by adding the exact value of the boot-up time and thus getting the exact time and date values are performed in this study. Since handling lots of hours, minutes and second values is hard and not logical when considering they are not totally different from each other in habit mining, hour is transformed into one of the four time slots.

5.2.4. Identifying/ Removing Outliers

It is known that modern mobile devices get the related local time from the carrier. However it has been observed that sometimes the mobile device cannot obtain this data straight away and instead it uses its default time and date value for a short time. This issue causes jumps up to 1970s. As stated previously, time value gathered by the application that collects data is based on the boot-up time of the device. In the case that boot-up time experiences such a jump in boot-up time, relevant times of all the interactions become outliers. Therefore interactions and values between this time of boot-up time and the following boot-up time are omitted for better results.

5.2.5. Elimination of Redundant Data

Recent apps that have been recorded by Android, is scanned in predefined 5-minute-intervals by the device analyzer. Together with the new interactions, the interactions recorded in the previous scanning session may be included in the current scanning session. Detecting which apps are genuinely newly interacted after previous scanning session is required in order to prevent the system to learn redundant knowledge. For this reason apps added on top of the previous list is used. Also, applications that were listed in the previous list but now have a new ranking between the other apps from the previous list are used. This situation can be best described with a scenario.

RECENT APPLICATIONS SCENARIO

Consider the first two application scans on a users' mobile device as following.

Table 3: First App Scan

Ranking	Application
app|recent|0	Facebook
app|recent|1	Gmail
app|recent|2	Chrome
app|recent|3	Angry Birds

Table 4: Following App Scan

Ranking	Application
app|recent|0	SoundHound
app|recent|1	Gmail
app|recent|2	Facebook
app|recent|3	Chrome
app|recent|4	Angry Birds

It is obviously seen that Angry Birds, Chrome and Facebook are coming from the previous scanning session. Since the ranking of Gmail gets higher it can be understood that the user has interacted with this interaction after the previous scanning session. And also SoundHound does not exist in the previous session list, this means that it is also a new interaction.

5.3. Training and Testing Phases

We have used data belonging to four users in our referred dataset. Each users mobile device type, the total duration of recordings and the number of applications installed on their mobile devices are summarized in the following table.

Table 5: Summarized Training and Testing Data

ser	Data source	Total Tracking Duration	Number of Applications Installed
User 1	Nexus 4	≅ 10 months	148
User 2	GT-I9100	≅ 14 months	89
User 3	Nexus 4	≅ 8 months	97
User 4	Nexus One	≅ 10 months	96

5.4. Parameter Tuning

Before starting the training phase parameters that needs to be tuned are minimum support and confidents thresholds that are used in rule induction phase of the approach proposed in the study. GCPM algorithm that was inspired from Cao et al. (2010), proposed setting minimum support threshold to two in the case when the participants of their evaluation had been tracked for one month. Therefore we tuned this parameter directly proportional with the number of months that user had been tracked. They proposed setting minimum threshold value for confidence as 0.5 because they argued that low confidence values is causing difficulties in distinguishing noisy data from association rules. However, setting this value to 0.25 increased the success of our study by 92% when compared to test results in which minimum confidence threshold is set to 0.5.

5.5. Definition of the Training Data

One of the studies that are implemented by carrying the similar aim to the scope of this study, is the study by Kamisaka et al. (2009). They have stated in their study that for sufficient prediction one month was enough.
Also Kurihara, Moriyama & Numao (2013) have based their study to collect data from a small PC and GPS unit for a duration of one month. In parallel with the studies found in literature, we have trained our system with training dataset that covers one month of user interactions.

5.6. Performance Metrics

The definition of the parameters used in evaluation of this study can be described as the following.

DEFINITIONS

- A *context* C_i is defined as the contextual feature-value pairs such as
  - $C_i = \{(x_1 : v_1), (x_2 : v_2), \dots, (x_l : v_l)\}$
    Where x_i denotes context features and v_i denotes values of these features.
- CS represents a *case* which is a tuple of context and the interaction preferred by the user and represented as
  - $CS_i = \langle C_i, I_i \rangle$
    Where C_i denotes context and I_i denotes the corresponding interaction in this case.
- P is *predicted interactions list* produced for a case and formulized as
  - $P = \{I_{p_1}, I_{p_2}, \dots, I_{p_5}\}$
    Where I_{p_i} denotes the predicted interactions.
- *Domain of applications* is represented as
  - $D = \{I_{d_1}, I_{d_2}, \dots, I_{d_n}\}$
    Where I_{d_i} denotes each of the interactions of the related domain.
  1. where $1 \leq n \leq \alpha_t$
  2. and
     $\alpha_t = \{9, 16, 32, \text{number of apps installed in the user device}\}$
- If a case record R = $\boldsymbol{CS_1 CS_2} \dots \boldsymbol{CS_N}$ are being tested in testing phase, performance of the approach is calculated according to the following performance metric.

PERFORMANCE METRIC

$$Performance = \frac{\sum_{k=1}^{N} x_k}{N}$$

$$where\ x_k = \begin{cases} 1, if\ P \supset I_k \\ 0, otherwise \end{cases}$$

$$\text{and } CS_k = < C_{i_k},\ I_k >$$

If I_k is available in the prediction list, the performance is increased by 1. This is performed for all of the cases.

5.8. Results and Discussion

In this section, results obtained after testing phase are displayed and discussed in detail.

5.8.1. Results

This section includes performances of approaches for each of the four subjects and distribution of these performances based on ranking is provided accordingly. In this section three approaches are evaluated. The first one is 2 Context Dimensional Hybrid Habit Mining Approach (2D-HHMA) which uses time and location as context. The second one is Complete Hybrid Habit Mining Approach (C-HHMA) which utilizes all context dimensions that are taken into account in this study. In C-HHMA, only complete match is accepted in retrieval phase. The last approach is Partial Hybrid Habit Mining Approach (P-HHMA), which is the approach proposed in this study. In P-HHMA, 10 different context dimensions are taken into account as it is done in C-HHMA. In P-HHMA, in case of not finding the exact case in the case base, only the match in time and location is determined as the condition of similarity.

5.8.1.1. Performances of Approaches

The following section includes the test results of the users in an easy-to-read graphical format. In these graphics, blue lines display the performances of approaches when interaction domain is limited to 9 and user interaction is predicted in top 5. Other lines indicate the performance values for various other candidate application counts (16, 32, unlimited respectively).

a) *Subject 1*

In the previous section of this chapter list of user profiles is provided. As described in this list, Subject 1 has 148 applications on his mobile device. Maximum number of test cases for this user is 20156.

Test results for Subject 1 are summarized in the Figure 18.

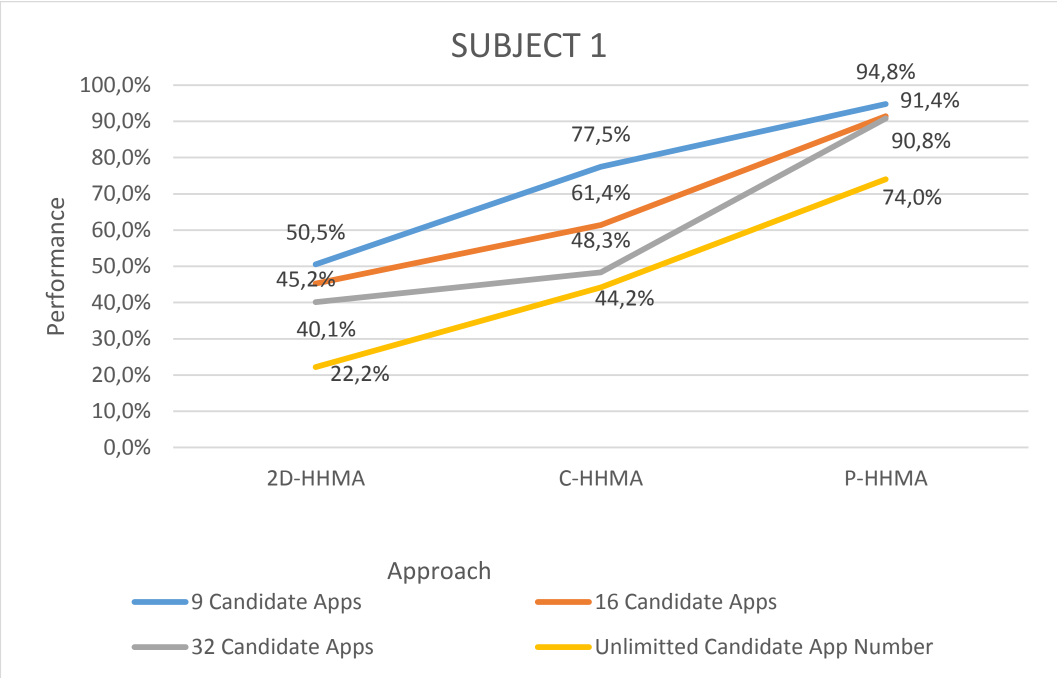


Figure 18: Test Results for Subject 1

b) Subject 2

Subject 2 has 89 applications on his mobile device. Maximum number of test cases for this user is 9098.

Test results for Subject 2 are summarized in the Figure 19.

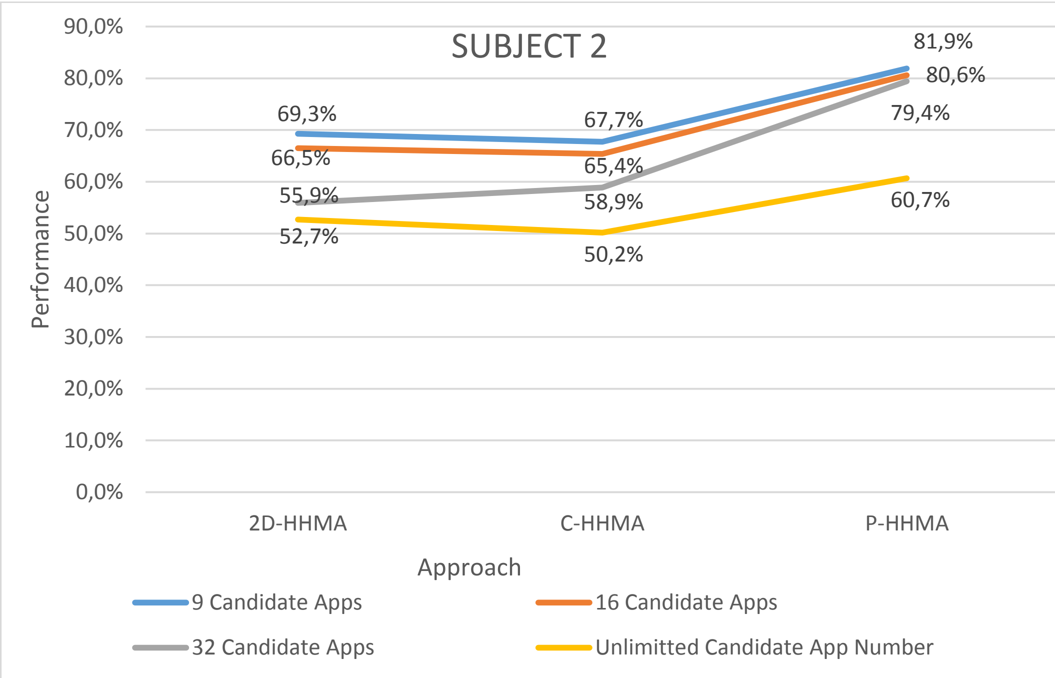


Figure 19: Test Results for Subject 2

c) Subject 3

Subject 3 has 97 applications on his mobile device. Maximum number of test cases for this user is 12639.
Test results for Subject 3 are summarized in the Figure 20.

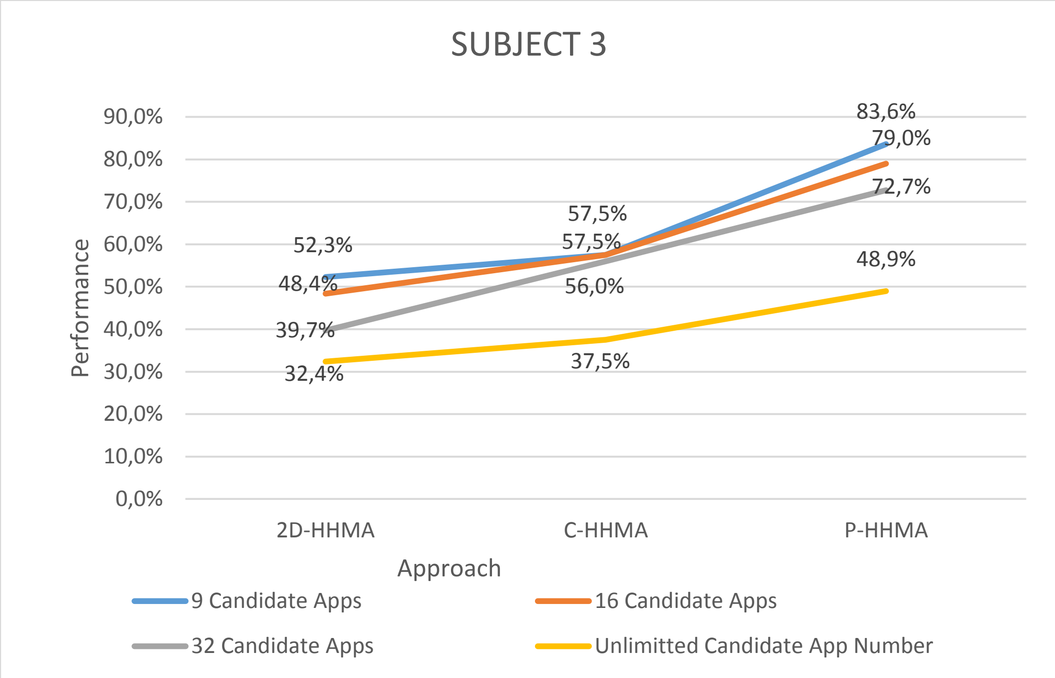


Figure 20: Test Results for Subject 3

d) Subject 4

Subject 4 has 96 applications on his mobile device. Maximum number of test cases for this user is 5950.
Test results for Subject 4 are summarized in the Figure 21.

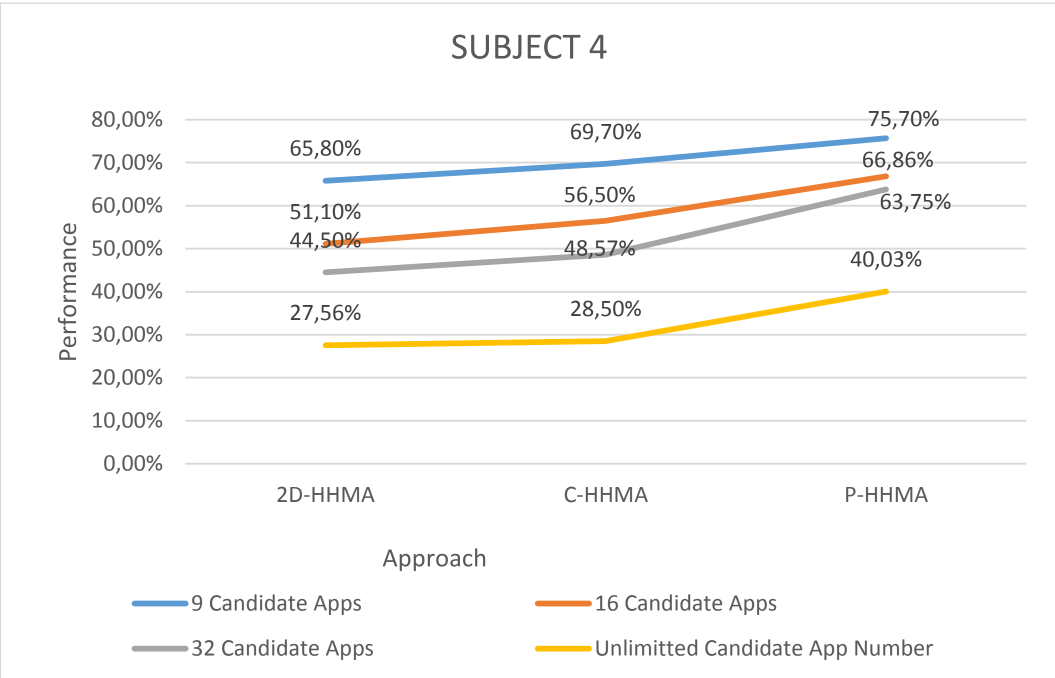
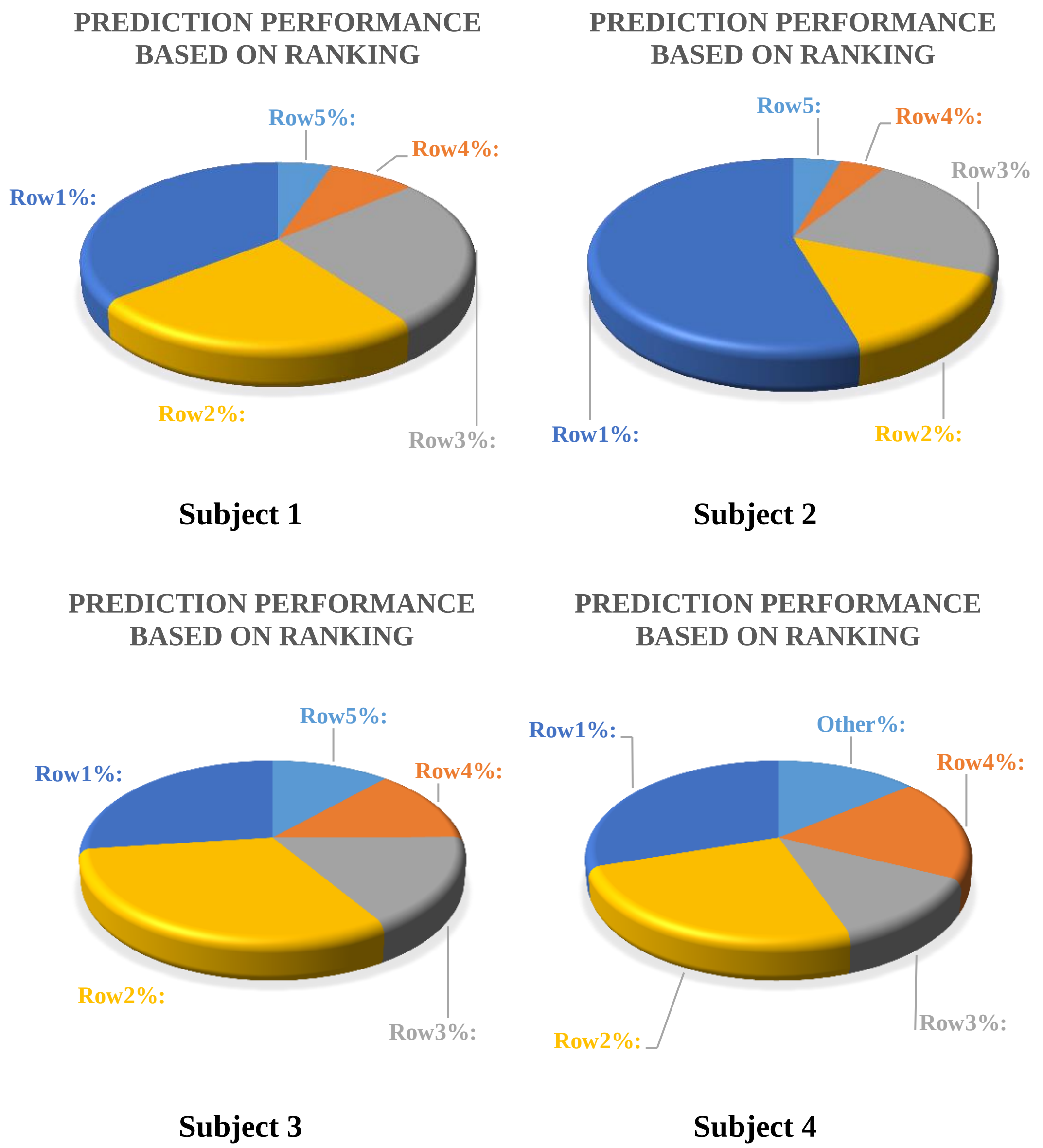


Figure 21: Test Results for Subject 4

5.8.1.2. Prediction Performance Based on Ranking

In the previous chapter, performances of approaches by using interaction domains containing different number of elements is shown. In Figure 22, distribution of performances of P-HHMA (with an interaction domain consisting of 9 elements) based on ranking for all users are displayed.



Subject 1 Subject 2

Subject 3 Subject 4

Figure 22: Prediction Performance Based on Ranking for All Subjects

5.8.2. Comparison with Previous Studies

Studies of Kamisaka et al. (2009) and Lee et al. (2011) do not have comparable metrics.

Cao et al. (2010) performed an evaluation that is based on some comments of their test subjects. Although they do not have metrics that allows direct comparison with this studies outcomes, their technique inspired this study and in hybrid approach their GCPM algorithm is used.

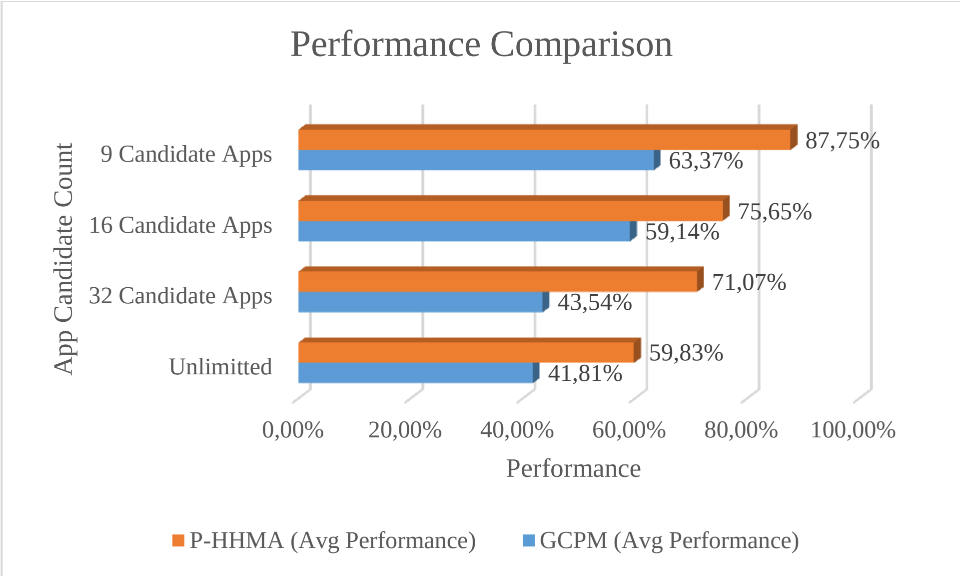


Figure 23: Performance Comparison with Different App Candidate Counts

The figure given above is an overview of the performances of GCPM and P-HHMA in different candidate application domains. Performances shown in figure reflect the average performances obtained for each test subject. For all cases, hybrid approach outstands. This is because even if GCPM can extract frequent behaviours, in real life, users are not so expectable. Therefore, specific non-frequent observed behaviours of users may be small but cannot be ignored.
It can be understood from the results that CBR compensates the drawbacks of generalization of GCPM.

Other approaches in the literature that share the same goal with our study are done by Cao et al. (2010) and Kurihara et al. (2013) as mentioned before. Some key-points of their evaluations are shown in Table 6.

Table 6: Key-points of evaluations of previous studies

# 1	Study by Shin et al. (2012)	✓ 23 Users for just over a month ✓ Average prediction accuracy recorded as 85% with 7 app candidates out of 32 applications.
# 2	Study by Kurihara et al. (2013)	✓ Their performance metric is measured according to whether preferred application is in top five out of nine items or not. ✓ Among five different subjects their recorded success rate is 53.8%, 72.9%, 50.2%, 66%, 65.4%

Comparison of hybrid approach proposed in this study (P-HHMA) with the mentioned previous studies is summarized in the following table.

Table 7: Comparison of P-HHMA and Other Techniques

Approach	**Performance metric**	**Accuracy**
P-HHMA	5/9	84.0%
# 2	5/9	61.7%
P-HHMA	5/32	76.7%
# 1	5/32	65-75%

P-HHMA achieves 84% when performance metric is predicting the most likely interactions in the top five out of nine whereas second study shown in the table (Shin et al. 2012) achieved 61.7% as an average of five test subjects. Evaluation results of P-HHMA is very close to study by Kurihara et al. (2013) when the same metric is used. On the other hand, this study which is symbolized as # 1 in the table, has records that are recorded in a short period of time. Since these kind of systems are learning systems, being tested after a shorter period of training effects the result. The exact testing periods of the subjects are not clearly stated in this study. Furthermore, different dataset may lead researchers into obtain different results. Stating a definite judgment is not possible in such a condition. But still, the obtained results can be considered as encouraging.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1. Conclusion

In this study, the primary aim is to propose a system that recognizes the situation of the user and behaves accordingly to make life of the user easier by thinking the way the user tends to think. This carries a big importance by means of saving time that is one of the core elements of life. The system is capable of being integrated to a dynamic user interface, specialized reminders, multi-functional recommender systems, and even very smart and complex personal assistants useful in many aspects. The proposed system contains a hybridization of CBR and GCPM that is derived from traditional association rule mining algorithms. Both of these techniques have upsides and downsides. GCPM is an efficient and effective approach to be used in behavior pattern mining. It derives the associations between user interaction records and corresponding contexts and detects the associations that are encountered frequently. However, detecting most frequently used applications is not enough if the subject is detecting all habits of the users to ease their life. Under certain circumstances such as specific contexts that do not frequently come across to the users, they may have other behavior patterns even more urgent and important such as catching a bus by checking its schedule. In such situations, CBR is very successful since it can retain cases even without any indication about their frequency. On the other hand, it can be considered as a high computational power and storage size required approach when it comes to utilizing it in a mobile device. To utilize this approach by using a remote server may result in not being able to track users at all times that they may need. Therefore, combing these two approaches provides an effective solution to the domain of mining mobile user habits.

Evaluation of the proposed model is done by using real-life dataset gathered by using DeviceAnalyzer. Four people in the set are kept track of for varying durations from approximately 8 months to 14 months. Results can be considered as encouraging when compared to previous studies in this domain.
The following conclusions are drawn in the study:

- GCPM is appropriate to use in detecting frequently used habits of the users.
- Real-word records of users however, are not so generalizable and generalization causes the loss of specific knowledge.
- Evaluation results of GCPM are not sufficient at all.
- CBR is used to compensate the mentioned downside of the GCPM.

- In study context is considered to have the dimensions of type of the day (weekend or weekday), period of day (morning, afternoon, evening, night), battery level (1, 2, 3, 4, 5), charger (AC, USB, disconnected), activity (stationary, walking, in vehicle), connection (Mobile, Wi-Fi, Disconnected), SSID -if Wi-Fi connected-, location (Home, Work, Out), status of airplane mode (on, off)\ status of ringing mode (vibrate, silent, normal). Accepting time and location is essential to increase the performance of the study.
- Personalized predictions of the system achieved different but promising results for each of the user.
- Hybrid approaches performed better when compared to the performance of GCPM.
- Comparing hybrid approach to other previous studies after restricting the candidate application count is again encouraging.

6.2. Future Work

In this study, different context dimensions are taken into consideration. Based on previous studies in literature, time is taken as an essential value together with the location (Saleh & Masseglia, 2011). Even this assumption increased the performance of the approach for the test set we used, this may not be valid for all users such as a war correspondent who travels the world and do not care so much about the local time of the region. As a solution, allowing users to select the most important context dimensions to be tracked manually, is proposed in the study. But still, a dynamic mechanism to adapt the weights of context dimensions for each user and to back propagate in each decision of these users to provide a progressive learning can be provided.

Also, it is possible to add different context dimensions to detect their effect on mining habits. DCON ontology (Scerri et al., 2012) can be useful in this manner.

Also, as a further step, developing a fully functional personal assistant that is capable of recommending in personal daily life domain can be performed by integrating the proposed system in this study.

REFERENCE LIST

Abowd, G. D., Dey, A. K., Brown, P. J., Davies, N., Smith, M., & Steggles, P. 1999, Towards a better understanding of context and context-awareness. *Handheld and ubiquitous computing* (pp. 304-307). Springer Berlin Heidelberg.

Akoush, S., & Sameh, A. 2007, Mobile user movement prediction using bayesian learning for neural networks. In *Proceedings of the 2007 international conference on Wireless communications and mobile computing* (pp. 191-196). ACM.

Attard, J., Scerri, S., Rivera, I., & Handschuh, S. 2013, Ontology-based situation recognition for context-aware systems. In *Proceedings of the 9th International Conference on Semantic Systems* (pp. 113-120). ACM.

Baralis, E., Cagliero, L., Cerquitelli, T., Garza, P., & Marchetti, M. 2011, CAS-Mine: providing personalized services in context-aware applications by means of generalized rules. *Knowledge and information systems*, *28*(2), 283-310

Barrett, B, 2010, *Evolution of the mobile phone*, viewed 31 December 2014, <http://edition.cnn.com/2011/10/05/tech/gallery/mobile-phone-timeline/>

Beaver, I., & Dumoulin, J. 2013, Applying MapReduce to Learning User Preferences in Near Real-Time. In *Case-Based Reasoning Research and Development* (pp. 15-28). Springer Berlin Heidelberg

Bellotti, V., Begole, B., Chi, E. H., Ducheneaut, N., Fang, J., Isaacs, E., ... & Walendowski, A. 2008, Activity-based serendipitous recommendations with the Magitti mobile leisure guide. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1157-1166). ACM

Bridle, R., & McCreath, E. 2006, Inducing shortcuts on a mobile phone interface. In *Proceedings of the 11th international conference on Intelligent user interfaces* (pp. 327-329). ACM

Bruke, R. 2002, Hybrid recommender systems: survey and experiments. *User Model. User-Adapt. Interact.* 12(4), 331–370

Cao, H, Bao, T, Yang, Q, Chen, E, & Tian, J 2010, An effective approach for mining mobile user habits, *In Proceedings of the 19th ACM international conference on Information and knowledge management* (CIKM '10). ACM, New York, NY, USA, 1677-1680.

Chen, A. 2005, Context-aware collaborative filtering system: Predicting the user's preference in the ubiquitous computing environment. In *Location-and Context-Awareness* (pp. 244-253). Springer Berlin Heidelberg.

Cooley, R., Mobasher, B., & Srivastava, J. 1999, Data preparation for mining World Wide Web browsing patterns. *Knowledge and Information Systems, 1*(1), 5-32.

Dey K. A. 2001. Understanding and Using Context. *Personal Ubiquitous Comput.* 5, 1, 4-7

Eagle, N., & Pentland, A. 2006, Reality mining: sensing complex social systems. *Personal and ubiquitous computing, 10*(4), 255-268.

Falaki, H., Mahajan, R., Kandula, S., Lymberopoulos, D., Govindan, R., & Estrin, D. 2010, Diversity in smartphone usage. In *Proceedings of the 8th international conference on Mobile systems, applications, and services* (pp. 179-194). ACM.

Funf Open Sensing Framework, *funf*, viewed 10 January 2015, < http://www.funf.org>

Gartner 2013, Gartner Says Worldwide PC, Tablet and Mobile Phone Combined Shipments to Reach 2.4 Billion Units in 2013, *Gartner Announcements*, Viewed 31 December 2014, <http://www.gartner.com/newsroom/id/2408515>

Gemmis, M., Iaquinta, L., Lops, P., Musto, C., Narducci, F., & Semeraro, G. 2011ö Learning preference models in recommender systems. In *Preference Learning* (pp. 387-407). Springer Berlin Heidelberg.

Girardello, A., & Michahelles, F. 2010, AppAware: Which mobile applications are hot?. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services* (pp. 431-434). ACM

Hintze, D., Findling, R. D., Scholz, S., & Mayrhofer, R. 2014 Mobile Device Usage Characteristics: The Effect of Context and Form Factor on Locked and Unlocked Usage.

Icev, A. (2003). *Distance-enhanced association rules for gene expression.* BIOKDD'03, in conjunction with ACM SIGKDD.

Ishio, T., Date, H., Miyake, T., & Inoue, K. 2008, Mining coding patterns to detect crosscutting concerns in Java programs. WCRE '08: Proceedings of the 2008 15th Working Conference on Reverse Engineering (pp. 123-132). Washington, DC: IEEE Computer Society.

Jannach, D., & Hegelich, K. 2009, A case study on the effectiveness of recommendations in the mobile internet. In *Proceedings of the third ACM conference on Recommender systems* (pp. 205-208). ACM

Karatzoglou, A., Baltrunas, L., Church, K., & Böhmer, M. 2012, Climbing the app wall: enabling mobile app discovery through context-aware recommendations. In *Proceedings of the 21st ACM international conference on Information and knowledge management* (pp. 2527-2530). ACM

Kamisaka, D., Muramatsu, S., Yokoyama, H., & Iwamoto, T. 2009, Operation prediction for context-aware user interfaces of mobile phones. In Applications and
the Internet, 2009. SAINT'09. Ninth Annual International Symposium on (pp. 16-22). IEEE

Krulwich, B. 1997, Lifestyle finder: Intelligent user profiling using large-scale demographic data. *AI magazine*, *18*(2), 37.

Kurihara, S., Moriyama, K., & Numao, M. 2013, Context-aware application prediction and recommendation in mobile devices. In *Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2013 IEEE/WIC/ACM International Joint Conferences on* (Vol. 1, pp. 494-500). IEEE.

Lee, H., Choi, Y. S., & Kim, Y. J. 2011, An adaptive user interface based on spatiotemporal structure learning. Communications Magazine, IEEE, 49(6), 118-124

Lee, J. S., & Lee, J. C. (2007). Context awareness by case-based reasoning in a music recommendation system. In *Ubiquitous Computing Systems* (pp. 45-58). Springer Berlin Heidelberg

Li, X, Cao, H, Chen, E, Xiong, H & Tian, J 2012, BP-growth: Searching Strategies for Efficient Behavior Pattern Mining, *Mobile Data Management (MDM), 2012 IEEE 13th International Conference on* , vol., no., pp.238,247.

Liu, D. R., Ke, C. K., & Wu, M. Y. 2008, Context-based knowledge support for problem-solving by rule-inference and case-based reasoning. In *Machine Learning and Cybernetics, 2008 International Conference on* (Vol. 6, pp. 3205-3210). IEEE

Mandl, M., Felfernig, A., Teppan, E., & Schubert, M. 2011, Consumer decision making in knowledge-based recommendation. *Journal of Intelligent Information Systems*, *37*(1), 1-22.

Nicolas, J. A., Herengt, G., & Albuisson, E. 2004, Sequential pattern mining and classification of patient path. *MEDINFO 2004: Proceedings Of The 11th World Congress On Medical Informatics.*

Porcel, C., Tejeda-Lorente, A., Martínez, M. A., & Herrera-Viedma, E. 2012, A hybrid recommender system for the selective dissemination of research resources in a technology transfer office. *Information Sciences*, *184*(1), 1-19.

Prentzas, J, & Ioannis, H 2007, Categorizing approaches combining rule-based and case-based reasoning, *Expert Systems*: 97-122.

Rawassizadeh, R., Tomitsch, M., Wac, K., & Tjoa, A. M. 2013, UbiqLog: a generic mobile phone-based life-log framework. *Personal and ubiquitous computing*, 17(4), 621-637.

Saleh, B., & Masseglia, F. 2011, Discovering frequent behaviors: time is an essential element of the context. *Knowledge and information systems*, *28*(2), 311-331.

Scerri, S., Attard, J., Rivera, I., Valla, M., & Handschuh, S. 2012, Dcon: Interoperable context representation for pervasive environments. In *In Proceedings of the Activity Context Representation Workshop at AAAI* (Vol. 2012).

Schiaffino, S. N., & Amandi, A. 2000, User profiling with Case-Based Reasoning and Bayesian Networks. In *IBERAMIA-SBIA 2000 open discussion track* (pp. 12-21).

Schilit, B. N., & Theimer, M. M. 1994, Disseminating active map information to mobile hosts. *Network, IEEE*, *8*(5), 22-32.

Shin, C., Hong, J. H., & Dey, A. K. 2012, Understanding and prediction of mobile application usage for smart phones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (pp. 173-182). ACM

Tarasewich, P. 2003, Designing mobile commerce applications. *Communications of the ACM*, *46*(12), 57-60.

University of Oulu, AWARE Android Mobile Context Instrumentation Framework, *University of Oulu*, viewed 10 January 2015, < http://www.awareframework.com>

Vu, T. H. N., Ryu, K. H., & Park, N. 2009, A method for predicting future location of mobile user for location-based services system. *Computers & Industrial Engineering*, *57*(1), 91-105.

Wagner, D., Rice, A., & Beresford, A. 2013, Device Analyzer: Understanding smartphone usage, *10th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, Tokyo, Japan.

Worstal, T, 2013, *Happy Birthday To The Mobile Phone: 40 Years Old Today*, viewed 31 December 2014,

<http://www.forbes.com/sites/timworstall/2013/04/03/happy-birthday-to-the-mobile-phone-40-years-old-today/>

Yan, B., & Chen, G. 2011, AppJoy: personalized mobile application discovery. In *Proceedings of the 9th international conference on Mobile systems, applications, and services* (pp. 113-126). ACM.

Zeng, C., XING, C. X., & Zhou, L. Z. 2003, A Personalized Search Algorithm by Using Content-Based Filtering [J]. *Journal of Software*, *5*, 018.

Zhu, H., Chen, E., Yu, K., Cao, H., Xiong, H., & Tian, J 2012, Mining Personal Context-Aware Preferences for Mobile Users. In *ICDM* (Vol. 12, pp. 1212-1217)

TEZ FOTOKOPİ İZİN FORMU

ENSTİTÜ

Fen Bilimleri Enstitüsü ☐

Sosyal Bilimler Enstitüsü ☐

Uygulamalı Matematik Enstitüsü ☐

Enformatik Enstitüsü ☒

Deniz Bilimleri Enstitüsü ☐

YAZARIN

Soyadı : BAYRAM

Adı : GAMZE

Bölümü : BİLİŞİM SİSTEMLERİ

TEZİN ADI (İngilizce) : A CONTEXT-AWARE APPLICATION RECOMMENDATION SYSTEM FOR MOBILE DEVICE USERS

TEZİN TÜRÜ : Yüksek Lisans ☒ Doktora ☐

1. Tezimin tamamından kaynak gösterilmek şartıyla fotokopi alınabilir. ☐
2. Tezimin içindekiler sayfası, özet, indeks sayfalarından ve/veya bir bölümünden kaynak gösterilmek şartıyla fotokopi alınabilir. ☒
3. Tezimden bir (1) yıl süreyle fotokopi alınamaz. ☐

Yazarın imzası .... Tarih ....