# SUGGEST ME A MOVIE: A MULTI-CLIENT MOVIE RECOMMENDATION APPLICATION ON FACEBOOK

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

BY

SEDA ÇAKIROĞLU

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
COMPUTER ENGINEERING

JUNE 2010

Approval of the thesis:

**SUGGEST ME A MOVIE: A MULTI-CLIENT MOVIE RECOMMENDATION APPLICATION ON FACEBOOK**

submitted by **SEDA ÇAKIROĞLU** in partial fulfillment of the requirements for the degree of **Master of Science in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences** ________________

Prof. Dr. Adnan Yazıcı
Head of Department, **Computer Engineering** ________________

Dr. Ayşenur Birtürk
Supervisor, **Computer Engineering Department, METU** ________________

**Examining Committee Members:**

Prof. Dr. Mehmet R. Tolun
Computer Engineering Department, Çankaya University ________________

Dr. Ayşenur Birtürk
Computer Engineering Department, METU ________________

Assoc.Prof. Dr. Nihan Kesim Çiçekli
Computer Engineering Department, METU ________________

Assoc. Prof. Dr. Ferda Nur Alpaslan
Computer Engineering Department, METU ________________

Asst. Prof. Dr. Pinar Senkul
Computer Engineering Department, METU ________________

**Date:** ________________

**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: SEDA ÇAKIROĞLU

Signature :

# ABSTRACT

SUGGEST ME A MOVIE: A MULTI-CLIENT MOVIE RECOMMENDATION APPLICATION ON FACEBOOK

Çakıroğlu, Seda
M.S., Department of Computer Engineering
Supervisor : Dr. Ayşenur Birtürk

June 2010, 87 pages

In this study, an online movie recommendation engine that serves on Facebook is developed in order to evaluate social circle effects on user preferences in a trust-based environment. Instead of using single-user profiles in the social environment identification process, virtual group profiles that present common tastes of the social environments, are formed to achieve a successful social circle analysis and innovative suggestions. Recommendations are generated based on similar social circles and based on social circles of similar users separately and their results are evaluated. Pure collaborative filtering is applied to emphasize the influence of social environment characteristics.

Keywords: Recommendation, collaboration filtering, social circle, facebook, virtual group profiling, movie, web application, social network

# ÖZ

SUGGEST ME A MOVIE: FACEBOOK'TA ÇOK KULLANICILI FILM ÖNERI UYGULAMASI

Çakıroğlu, Seda
Yüksek Lisans, Bilgisayar Mühendisliği
Tez Yöneticisi : Dr. Ayşenur Birtürk

Haziran 2010, 87 sayfa

Bu çalışmada, Facebook uygulaması olarak ulaşılabilen bir film öneri sistemi tasarlandı. Bu sistemden gelen bilgilerin analizi ile sosyal çevrenin kişisel zevklere olan etkisi anlaşılmaya çalışıldı. Daha verimli ve faydalı bir sosyal çevre etkisi analizi yapmak amacıyla, kişisel profiller yerine, kullanıcıların arkadaş çevrelerinin ortak zevk ve tatlarını gösteren sanal profiller oluşturuldu. Daha yenilikçi ve doğru film önerileri yapabilmek için bu oluşturulan sanal profillerin uyumluluğuna göre öneriler yapılmaya çalışıldı. Esas olarak, benzer sosyal çevreye göre ve benzer kullanıcıların sosyal çevrelerine göre olmak üzere iki farklı yöntemle film önerileri oluşturuldu ve öneriler değerlendirildi. Daha doğru bir analiz yapabilmek amacıyla öneri belirlenmesinde sadece iş birlikçi filtreleme kullanıldı.

Anahtar Kelimeler: Öneri Sistemleri, iş birlikçi filtreleme, sosyal çevre, facebook, grup profilleme, film, web uygulaması, sosyal ağ

*To my lovely father*

ACKNOWLEDGMENTS

My sincere thanks to my advisor Dr. Ayşenur Birtürk for her extensive knowledge and support that she gave me through all this study.

I want to thank to TUBITAK for the support through my academical life.

My special thanks to my lovely family, everything that I have, Fahriye Çakıroğlu, Hasan Çakıroğlu, Eda Çakıroğlu and Arzu Uçarlı. Thanks to Mert Akdemir for his belief in me and all the innovative ideas that he supplied. Sincere thanks to Goncagül Demirdizen for all the support that she supplied.

Lastly thanks to the anonymous reviewers of Webist'10, ACM RecSys'09 and AAAI'10 and ISCIS'10 conferences for their useful feedbacks.

# TABLE OF CONTENTS

# LIST OF TABLES

TABLES

# LIST OF FIGURES

FIGURES

# CHAPTER 1

# INTRODUCTION

We are now in a new information world which is a large network that is in interaction with us in every second of our lives. In such a huge network, it is very hard to identify where and how the needs can be fulfilled accurately. Most of the people are lost in this huge information batch.

Information filtering systems gain their importance at the point of helping people to find their ways in this network. They are like self-assistants who can concern about personal needs and requirements and capable of finding the correct way to reach the required data. Recommendation systems are one of the most important types of the information filtering systems. Distinctively from information filtering systems, the main concern of recommendation systems are our likes and dislikes.

In recommendation systems, the collected information from all over the world is served to us by applying some filtering processes. By this elimination strategy, losing our ways in this huge raw data is prevented. Recommendation systems serve as personal specific pleasure agents for the users. Every agent is responsible for knowing its own master and filtering the data considering the preferences and tastes of the master. With this personal specific guidance methodology the usage of recommendation engines supply us with the advantage of owning our personal guides who only think for our fancy.

## 1.1 Problem Description

Considering the power of the recommendation systems and their success in the market, the related search area is growing very rapidly. In industry and academia, searches are done in order to identify the most efficient and effective way of recommending the best matching items to the users. Through these researches, different areas of science such as Artificial Intelligence, Natural Language Processing and Sociology are considered with the aim of increasing the success of the recommendations. Several workshops and conferences about Recommendation Systems are performed to share the existing know-how of how to recommend more accurate and efficient way.

Since all the recommendation algorithms are rooted from the same ideology which is suggesting the most suitable items to the users, as a side effect, most of the algorithms are very similar to each other. In most of the conditions, the generated recommendations of these algorithms are common and even sometimes they are the same. This fact creates a new problem which is accessing the same suggestions every time although different recommendation engines are used. As an instance, for a user who is familiar with the recommendation systems, experiencing the same suggestions in every recommendation engine is boring and meaningless. Considering this case it is clear to say that new aspects are required for the innovation of the algorithms and their recommendations.

Moreover, at the end of the trust researches on the recommendation system area, it has been identified that people give more importance to their friends' suggestions rather than the ones of the recommendation systems [37]. Although the social circle effect on the personal preferences are searched in several studies, in the current recommendation engines, there is the lack of considering social circles as a single source of suggestions regardless from the users' personal preferences. It is required to perform more detailed searches for the usage of social circle's suggestions in the recommendation generations.

Although social networks are used as a source of information for the personal preferences, the usage of this data is very limited in the current recommendation algorithms. Most commonly the suggestions are generated considering the close related friend information on the social networks. However it is required to use this data in a more common and extended way. The generic profiles that represent the common tastes of the social circles of the users should be created with the aim of increasing the success of the algorithms. By this way, the commonly shared preferences in the social circle profiles can be emphasized. The possibility of generation of accurate suggestions can be increased.

## 1.2 Aim of The Study

In this study, it is aimed to identify the importance of the social circle effect on human preferences in the recommendation area. A movie recommendation engine has been studied on social network, Facebook, with the intention of generating innovative suggestions to the users based on the data that is gathered from the social circles. Two basic algorithms are studied for recommendation generations which are 'Social Circles of Similar Users' and 'Similar Social Circles'. Both of the algorithms are tried to be evaluated statistically considering the accuracy and innovation of their suggestions.

In the 'Social Circles of Similar Users' algorithm, the identification of the similar users for every user is performed. Regarding to this similar user data, the suggestions are tried to be generated based on the preferences of the social circles of these similar users. In the 'Similar Social Circles' algorithm, it is aimed to generate suggestions regardless from the personal preferences. After identification of the similar social circles of users' social circles, the recommendations are generated considering the preferences of these similar social circles.

## 1.3 Outline of The Thesis

This thesis consists of the following chapters:

**Chapter 2 - Recommendation Systems** gives general information about the information filtering and recommendation systems. The commonly used methodologies in the current recommendation algorithms are described. Additionally, the reason of the selection of Facebook as the development environment is stated.

**Chapter 3 - Examples of Information Filtering System** supplies a brief description of the popular recommendation systems that inspire this study.

**Chapter 4 - Methodology** describes the recommendation algorithms and the characteristics of the domain in which the study is performed.

**Chapter 5 - System Architecture** presents the implementation details of the recommendation engine. The detailed design architecture of the application and the recommendation agents are supplied.

**Chapter 6 - Evaluation** stated the evaluation methodologies and metrics of the studied algorithms. The results of the evaluation phase are shared and the comparison between two algorithms are performed.

**Chapter 7 - Discussion** describes the problems that have been experienced through the design phase of the study and corresponding solutions that are applied.

**Chapter 8 - Future Work** suggests some possible improvement aspects which can increase the success of the studied recommendation engine.

**Chapter 9 - Conclusion** provides a general overview of this thesis and draws the conclusions of the study.

# CHAPTER 2

# RECOMMENDATION SYSTEMS

## 2.1 Information Filtering Methodologies

Currently on the web you can access every kind of data if you know where you can find it. Be sure that the music that you are searching for was uploaded before and the answer that you are looking for has been already given. Considering this extensive knowledge on the web, the researchers have started to look for an answer for a different and more complicated problem which is how we can find. Taking into consideration that there is very huge data batch on the web, it is required to suggest a methodology which helps to easy access to what is required [1].

With the aim of suggesting a way to filter the required data regarding to the users' requirements, two basic information processing methodologies are discussed, which are 'Information Retrieval' and 'Information Filtering'.

### 2.1.1 Information Retrieval

Information retrieval engines are the search engines that are used very commonly on the web. The aim of these search engines is accessing specific data in the shortest time and the most accurate way. The correctness and efficiency of the accessed data is the most important criterions for the success of these information retrieval engines. They basically aim to perform temporary search queries to find an answer for a specific question that is asked by the user [2].

As example, the most commonly used information search engine, Goggle [3] can be categorized as an Information Retrieval engine. It uses a very similar algorithm with Kleinberg's clustering and ratings algorithm [4] to return answers to the search queries of the users. In this algorithm, it is tried to understand and analyze the rich source of information on the web considering text matching and usage rate of the web data. For an additional illustration of the methodology, IMDB [5], the one of the most popular movie search engines can be supplied as an example.

### 2.1.2 Information Filtering

In the Information Filtering methodology the aim is identifying user needs and requirements and notifying the users about the new activities based on this gathered profile data. The methodology tries to track new data and decides the parts that should be shared with the users based on the user requirements and preferences. Email notification systems can be shown as an example of Information Filtering engines. In such kind of engines, according to the subscribed categories, users are notified about new developments by emails.

By the searches that are done in this area, more advanced Information Filtering engines has been built which are capable of learning about the personality characteristics of the users. As an example in the Maes' agents [6] and Lieberman's Letizia [7], the filtering of the information is performed on the e-mails and news data regarding to personal requirements that are learned through the tracking of the web activities of the users. By this way, the initial elimination of the information on the web is performed based on the gained data about users' profiles before serving it to them. As another example of the sophisticated Information Filtering engines, Cohen's Ripper system can be indicated [8]. This system aims to filter emails using learning rule algorithms and term frequency analysis [9].

Since Information Filtering algorithms try to gather data about the user profiles, they are one of the most important key aspects for the recommendation algorithms. The same principles are applied in both of methodologies in order to access personal characteristic data of the users.

## 2.2 Recommendation Algorithms

The searches and studies on the market show that Recommender Systems are very popular and attracts the attention of the users. Especially considering commercial and practical results, it is identified that users prefer the personalized engines which can think at personal base instead of the engines that behave in the same way for every user [10]. Considering this huge data garbage on the web, users have started to search for the opportunities that help them to select what to view or listen. With the help of the recommendation systems, people find the chance of accessing pure and simple data that is filtered in a personal perspective.

In another point of view, the aim of recommender systems can be stated as creating a methodology that reorganizes data that is collected from entire population. Basically they try to use this reorganized data for the other people in a more beneficial and efficient way [11]. For an example if you do not know which wine you will drink, you can use the existing information on a wine recommender and you can find the best delicious wine for you using the know-how of other people and your personal preferences [12].

Recommendation Systems try to learn the required personal information about the users through implicit or explicit rating methodologies or applying both of them [13]. Explicit rating means asking users about their real opinions through survey or questionnaires whereas implicit rating means examining the inputs and assigning predicted rates based on a pre-decided scale matrices [14].

Although it is accurate to say that explicit rating supplies more accurate information rather than the implicit rating methodology, for most of the users it is very boring to answer unending questions and comment on the long list of the preferences. As a successful applicable example of the implicit feedback usage methodology for the content based filtering algorithm, Steven's search can be declared [15]. By this research, it has been proven that accessing correct information about the user profiles is possible by implicit rating methodology without disturbing the users.

For a recommender system, the most important next question after the identification of the user profiles is that how to decide what should be recommended. By stating the similarity of the preferences of the users or the content of the items, the suggestions are generated in a personal based way for every user. In most of the studied recommendation systems, two basic approaches are used for the generation of these recommendations. These methodologies are Content-Based Filtering and Collaborative Filtering approaches [16]. Commonly in the recommendation systems, one of the stated approaches or a hybrid approach that combines both of the methodologies are used [17] [18].

### 2.2.1 Content-Based Filtering

In the Content-Based Filtering methodology, as the main aspect, the items are tried to be categorized instead of the users [19]. The recommendations are generated considering the items and their contents. With the categorization of the items in the Content-Filtering approach, the recommendation engine tries to identify the weights of the ratings of the users on the item categorization features. The recommendations are generated from the unrated items for which the matching features have been highly rated by the users.

As an illustration of the Content-Based Filtering approach, movie recommendation systems can be considered. For the content categorization of a movie, the criterions like year, artists, genre and plot can be used. Multiple criterion usage supplies algorithms with a more accurate and complete categorization of the items for the recommendations. However it is also beneficial to use only one criterion in order to identify the impact of the parameter on the success of the recommendation algorithm. For example only publication year of a movie can be used as the content information in order to identify the impact of the situational environment when the movies were filmed [20].

Near these, as a new aspect of collaborative filtering methodologies, with the aim of increasing the recommendation's success, conversion of collaborative user models to content based user models are studied and applied in the recommendation engines. By this way it is aimed to enrich the data of content-based user profiles [21].

### 2.2.2 Collaborative Filtering

In Collaborative Filtering approach, the main focus is on the users. It tries to identify the profile of the users using Information Filtering techniques that are described before. The rooting idea is to find similar users with the aim of sharing the preferences and choices between them. Differently from the Content-Based Filtering approach, Collaborative Filtering does not think about the content information of the items.

The biggest advantage of Collaborative Filtering approach is that explicit content description is not required for applying this methodology. That means it is not needed to ask too many questions to the users about their preferences and tastes. By this approach, collected data from the other users can be reused for the recommendation generations. Regarding to this, the advantage of generating better recommendations to the users who are in cold-start phase is achieved in the collaborative filtering approach.

### 2.2.3 Well-Known Examples of Collaborative Filtering

At the beginning, in the firstly designed recommendation systems, it was required to collect data from the users explicitly [22] which created too much workload on the users. With the improvement of the algorithms and the enrichment of the user data sets, now the agents that require very limited information to find the similar users are designed. In these new sophisticated systems, the suggestions and predictions are generated based on very limited personal information [23].

Some of the well-known collaborative filtering recommender system studies can be declared as GroupLens, Ringo and Bellcore's Video Recommender. Konstan's recommendation engine, GroupLens generates the suggestions with the usage of very limited personal information and aims to filter news on web using the existing know-how of the similar users. It studying to filter the net news by applying collaborative filtering approach [24].

Ringo [25] can be referred as another successful instance of the collaborative filtering recommendation engines. In particular, Ringo is implemented as a networked system, which generates personalized recommendations by using the collaborative filtering principles and the content of the music albums and artists. It develops recommendations extremely fast with the usage of the analyzed audio data in the recommendation generation process. For a different domain recommendation engine, Bellcore [26] movie recommendation engine can be stated as an example which aims to recommend on the movie domain.

### 2.2.4 Bottlenecks of Collaborative Filtering

However in contrary to the advantages of the Collaborative Filtering approach, it is required to indicate that as a disadvantage, easy attacks by malicious insiders is very easy in this approach [27]. The attacks are basically performed by copying of the actual users' preferences on imitation users and changing the overall ratings of some of the items using these imitation users. As an example, a music recommender that suggests the same track to all of the users because of its higher rate can be illustrated. By manipulating the rating data of the track, this strategy can be made truth, which concludes several download of the track.

Since the Collaborative Filtering is a commonly used methodology on the commercial web sites, it is not a tolerable behavior. For overcoming this difficulty, friend of friend methodology is started to be searched [28] which discovers for the actual personality of the users in order to prevent the usage of data of the imitation users. Also as another approach, the methodology of creating trust matrices are studied by Massa and the success has been approved by the statistical results [29].

Moreover, as another disadvantage, in collaborative filtering recommender systems, it is not possible to obtain the mental model match [30] every time for every user. Under these circumstances, recommendations are tried to be generated from the most commonly liked items which is not accurate for everyone. In this condition, the usage of these highly rated items in the suggestions means referring the same suggestions for every user which is completely contrary to the personal specific suggestion ideology of the recommender systems.

## 2.3 Profile Analysis in Recommendation Systems

The social circle analysis is rooted from the idea that recognizing the persons rather than the users. For clarification, 'Recognizing The Person' can be explained as identifying person's preferences and tastes in an overall perspective considering the cultural and social effects which shape the personality. The aim of using this methodology can be stated as obtaining a trustful environment with actual users and knowing these users better.

As an important example, the person modeling idea has been studied through Interest Map methodology, which tries to identify the social personalities of the users by tracking the evidences that they leave on the web [31]. The study used the interest of music, books, television and all the other supplied information that was accessed through the social network site, MySpace. In fact for this study, the idea was rooted from the Social Data Mining approach which tracks the data of Usenet messages [32]. The study tries to create the opportunity of sharing actual preferences and opinions between people.

Interest Maps aims to analyze the patterns of interests considering social and cultural identities of the persons. By the investigation of co-occurring of the patterns, an interest map had been tried to be identified. Basically, map was created as identity versus interests of books, music and movies. Interest Maps have been studied in the recommendation area since 1975, by spreading activation over the network [33] with the consideration of the recommendation problem as a semantic context problem.

### 2.3.1 Social Circle Analysis

The Social Circle Analysis approach for generating recommendations is the main concern of this study. The idea that the persons who we care much are presenters of our social circle was firstly studied in the search of Donath and Boyd [34]. It is accurate to say that for every user, the activities that are pointed in their profiles and their connected friends are selected as a virtual representation of a social game character. This character is used in the recommendation generations without guaranteeing the reality of the personality.

Since network is used for making friends, dates and business connections, there is the possibility of obtaining misleading information in the case of using profile data without applying any filtering process. Therefore a more detailed analysis on the personalities of the users should be performed in order to achieve better recommendations.

The strategy of seeing a group of people as a single person and building a single memory structure on these groups has been already studied in the organizational area [35] [36]. In those studies, commonly it is aimed to create a shared know-how of the organization members. In this thesis study, the idea is extended by the consideration of the social circles of the users. Similarly a common memory is formed for a group of people in which family and friends are included. A virtual group profile is tried to be identified considering the social circles of the people.

Besides, it has been identified that the recommendations of friends are always preferable to the recommendations of the recommender engines [37]. It is clarified that users prefer their friends' suggestions considering their quality and usefulness of them. By the 'Suggest Me a Movie' application that has been studied through this thesis, it is aimed to generate recommendations considering the existing knowledge of the recommendations of the friends. For that reason, common virtual profile that presents the social circles of the users are formed.

## 2.4 Reason of Selecting Facebook as The Development Platform

The basic intention of this study can be stated as using the existing user data on the web in order to build a more trustable environment. For this aim, at the architectural design phase of the study, blogs and social networks were analyzed and the advantages that they will supply were searched.

Blogs are one of most rapidly expanding area which grows in exponentially [38]. Considering multi-domain data on the blogs, it is accurate to say that there are too many clues on the blogs to recognize someone. However considering the unstructured data collection way of blogs, it is decided use the social networks instead of them. In addition to this reason, regarding to the algorithm structure, it is also concluded that the first priority is accessing friend information of the users which is the main concern of the social networks.

As the next step, a social network list is generated with the aim of identifying a suitable development environment for this study. The options that are considered can be listed as, Orkut [39], MySpace [40] and Facebook [41]. These are the most popular social networks all over the world. Considering the study requirements and the current popularity of Facebook, it was decided to use Facebook as the development environment. In addition to the popularity advantage, since it supplies its developer with an advanced help structure, this environment seemed as the best choice for the implementation of the studied algorithms.

## 2.5 Explanation of The Generated Recommendations

The idea of sharing the reason of the suggestions are rooted from the intention of increasing transparency and trust of the recommendation engine. The studies showed that trust in recommender systems is one of the key properties that attracts attention of the users [42]. In these studies it is indicated that what the users really want is understanding the reason of the suggested recommendations and why they are suitable for them. The searches on reasoning and trust of recommender systems is extended by Herlocker by statistically proving that reasoning of the suggestions increases trust and acceptance rate of the users [43].

Regarding to these researches, on the ’Suggest Me a Movie’ profile page on Facebook, the recommendation methodologies of the application are shared with the users in a user friendly language. By this way, the basic principles of the recommendation identification process are stated. In addition to this, the algorithms that are used for the generation of the suggestions are shared with the users at the recommendation phase.

# CHAPTER 3

# EXAMPLES OF INFORMATION FILTERING SYSTEMS

Information Filtering systems are commonly used in several domains in order to suggest the best matching items to the users. Especially after the identification of the benefits of the information filtering systems in the marketing area, the web developers are started to use these filtering algorithms more commonly in their applications.

Since users like the idea of having a personal guide in the domains that they have interested in, the popularity of the information filtering algorithms are increasing exponentially. The users experience the chance of accessing what they really want only by supplying very limited information.

The popular domains that use information filtering can be listed as Music, Movie and Friendship sites. In this section, some of the popular engines that serve on web are stated and the main features of them are described. Near this their contribution in this thesis study are indicated.

## 3.1 Last.fm, Music Recommender Engine

Last.fm is one most the most popular music recommendation engine on web that has approximately 20 million users over 200 countries. It aims to recommend its users about music related issues considering personal tastes and preferences [44]. Last.fm serves its users through several platforms such as its internet station and plug-in applications for ipods, iphones and web.

After the account creation on the Last.fm, users are allowed to create their own radio stations, play lists, groups and events and share them between their friend lists. The users have the chance of being notified about events, festivals and concerts that they are interested in.

The identification of the users' likes and dislikes are performed based on the activities of the users such as listened tracks or assigned groups. Collaborative Filtering approach is used basically for the recommendation generations. The enrichment of the collaborative algorithm is done with the usage of 'social tags' which are predefined or free text tags that are available for tagging the items.


Figure 3.1: The Interface of Last.fm

In the figure 3.1, the general interface of the application is stated.

Last.fm uses collaborative filtering approach for generating the recommendations. The profile information of the users is supplied by tracking the actions of the users through web, internet radio station and plug-in applications. Near the tracked information, Last.fm enriches the profile data with the consideration of the comments that users supply to the recommended items.

For commenting on a recommended item, Last.fm supplies three options to its users, which are:

- Love this track
- Skip this track
- Ban this track

In addition to the supplied comments by the users, tags are one of the most important aspects of the Last.fm recommendation algorithm. With the intention of identifying users' likes and dislikes, tags are used for grouping tracks and artists. For applying the collaborative filtering approach, Last.fm tries to identify similar users and to suggest based on their existing choices considering the declared social tags.

Tagging mechanism in Last.fm is named as Social Tagging, which is rooted from the idea of sharing tags between the social circles of the users. Users are allowed to assign as many tags as they want to the items including genre tags and personal impression tags. According to the selected tags, the profile information of the users are enriched. Regarding to the statistical results, currently the most popular tags in Last.fm are rock, seen live, alternative and electronic.

Similar users which are called neighbors by the Last.fm, are started to be identified immediately after a user sign up the application [45]. With the data of five tracks that are commented by this user, the recommendation engine starts to search for the neighbors. In one week period, the neighbor list of the user is constructed. With the identification of the neighbors, recommendation generation process is started. The neighbor list is refreshed for all users in every week regularly. The enrichment of the neighbor set is performed with the data retrieved in the corresponding week [46].

Last.fm also serves as a social network on the web. It allows creating friendship connections between users. By this way, users have the opportunity of sharing the tracks or artists that they like with their own friends. Additionally, users also have the chance of defining new friend groups who share the same music taste.

As an improvement to the available collaboration filtering algorithms, Last.fm designed a strategy in order to prevent the dominant effect of the users who listen more music, at the similarity decision. By this enhancement, the information amount on the user profiles is excluded from the similar user analysis process.

## 3.2 FOAF Friend of a Friend Project, Friend Finder

Friend of a Friend (FOAF) Project was created as an experimental linked information methodology in 2000 with the aim of showing our place in the Web, and the Web's place in our world. It supplies with a RDF/XML syntax to create personal profiles on web including the information of the activities and friends. The aim of the FOAF project can be stated as allowing sharing the defined personal meta data between different network sites.

The main aspect of FOAF project is basically creating a simple way to share and use personal information on the web. FOAF tries to interconnect, merge and use the personal data and activities of the users between different social network sites. The merged and connected data is used for the identification of the connections which help to find the position on the web and access to the other people.

The personal raw data is collected from the social network sites and the data is related to each other considering the criterions like people places, organizations and documents. With the specified RDF/XML syntax, the FOAF project helps to define the meta data about people, their interests, relationships and activities [47].

FOAF uses RDF/XML syntax to indicate and share data between social networks. RDF is a new technology that is announced by World Wide Web Consortium (W3C). It is defined as Resource Description Framework which serves a methodology to describe things and relationship between these things.

In FOAF project, a simple RDF/XML vocabulary is served to the users. The library is used for defining the users considering their publications, employment details, group memberships

```
<rdf:RDF
     xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
     xmlns:foaf="http://xmlns.com/foaf/0.1/">
<foaf:Person>
     <foaf:mbox rdf:resource="mailto:simon@xmlhack.com" />
     <foaf:depiction rdf:resource="http://example.org/photos/edd+simon.jpg" />
</foaf:Person>

</rdf:RDF>
```

Figure 3.2: The RDF/XML code of Person Tag

```
<foaf:Person>
      <foaf:name>Edd Dumbill</foaf:name>
      <foaf:mbox rdf:resource="mailto:edd@xml.com" />
        ...
      <foaf:knows>
           <foaf:Person>
                <foaf:mbox rdf:resource="mailto:simon@xmlhack.com" />
                  <foaf:name>Simon St.Laurent</foaf:name>
           </foaf:Person>
      </foaf:knows>
</foaf:Person>
```

Figure 3.3: The RDF/XML code of Know Tag

and the declared interests. For example, in the RDF/XML definition in Figure 3.2, you can see an example of how the user personality is defined with the ’person’ tag. The FOAF project uses mail addresses as a personal unique identifier considering the fact that user name is not unique enough. In addition to these in the second RDF/XML (Figure 3.3), the friend relationship is indicated with the ’knows’ tag [48].

The motivation idea behind the FOAF project can be specified as sharing of the extensive data on one of the social network with the other social network sites. With the usage of the stated RDF/XML syntax, the data sharing is simple and applicable between different networks.

As another example of Friend of a Friend methodology, Jung study can be declared [59]. In Jung’s study, the personal preferences of users are tried to be identified considering the clues that they leave on a social network, MySpace. The study uses a data set which is constructed based on the data on MovieLens. The basic aim of the study is achieving the visualization of recommender algorithms on social networks. The study basically uses user similarity and friend of a friend methodology.

## 3.3 Facebook, Social Network and Friend Recommender

Facebook is a social networking web site that allows its users to communicate with their friends through the profile updates, messaging and application usage. It allows sharing information between users considering the privacy settings that is specified by every user himself/herself. Activity, video, audio, game and application sharing is allowed on Facebook. According to the latest researches Facebook is the most popular social network that is followed by MySpace. Currently it has more than 350 million active users.

Facebook was firstly created as a communication platform between Harvard University members by Mark Zuckerberg, Eduardo Saverin, Dustin Moskovitz and Chris Hughes. Although the application was created only for limited social area, it has been expanded its user profile firstly to Boston and then all over the world [52].


Figure 3.4: The Interface of Facebook

In the Figure 3.4, the general interface of the application is stated.

Near the stated activities, Facebook also serves as a friend recommender engine which aims to recommend possible friends to its users. In the rest of this section, considering the contribution of Facebook to the studied algorithms, only friend recommender engines on Facebook will be described.

Facebook performs friend recommendations with the aim of extending the social circle information of its users. There are two different Friend recommendation approaches that are used in Facebook for generating suggestions. These approaches are;

- People You May Know: Recommendation considering social circles on Facebook.
- Friend Suggester: Recommendation considering other web activities on different web applications.

In the 'People You May Know' approach, Facebook recommends based on your social circle information. The stated mutual connections while adding a new friend, for example 'went to school together', 'worked together', are used in the construction of your social circle graph. These mutual user connections in the constructed social graph are analyzed by Facebook recommendation engine and possible friends are referred to the users. The best suggestions are tried to be made by considering the closeness degree of the user to the indicated possible friends. For this degree, the distance in the graph representation is used basically.

Considering that in this approach the data that has been already supplied by the user is analyzed, no additional permissions is required for using this functionality. Facebook uses the user mutual connection data without asking further permission. Near this, it suggests actions that user can take to help friends that are new members become acclimated to the Facebook environment.

The other approach 'Friend Suggester' is a new feature that Facebook serves its users. Differently from the first strategy, this feature has taken friend recommendation strategy of Facebook to another level. It uses the data retrieved from the other social network sites and email accounts. As an example the contact list of your yahoo email address is served to you as a friend suggestion in Facebook by the usage of this approach. By this way, Facebook achieves to use multi-platform data about its users to generate recommendations.

However since the methodology uses the users' private data in other platforms, additional permission is required to activate this suggestion approach on your Facebook account. After the required permission is supplied, Facebook starts to recommend to you based on your data in other platforms.

## 3.4 Netflix, Movie Recommendation Engine

Netflix is a movie recommendation engine that serves on web with the aim of helping people to find the movies that they may like. Nowadays Netflix is one of the most popular recommendation engines in the movie domain with the advantage of its successful recommendation algorithm, the huge movie data set and its flexible application environment.

Netflix tries to recommend based on the tracked items and the comments that users supply. Near this personal data gathering methods, Netflix uses the data of the friends of the users in order to enrich the profile information. The tracked movies and videos in the list of the friends are considered in the recommendation generation process. A structure that is named as queue is used on Netflix, which includes the movies that it is ordered or flagged as 'I want to watch'. Users are allowed creating their own queue contents and share their queues between their friend lists.

In the Figure 3.5 the general interface of the Netflix web application is given. The flexible interface of the Netflix is one of the key aspects of the application that attracts the users' attention. Regularly in two week intervals, the interface of the application is changed and the opinions of the users are asked. By this interactive feedback mechanism users have the chance of selecting the best interface that they want to use.

Netflix uses collaborative filtering approach as the recommendation algorithm. It gathers data by analyzing the users' viewing habits and recommends considering their histories on the Netflix site. Netflix movie library is one of the largest movie data set that is used in the movie recommendation area. It stores and use 100.000 movies including their basic information such as title, cast, genre and plot. In this library near the classic movies information, new releases and independent movies are also stored and used in the recommendation algorithm. The library is open for all the researchers.

The collaborative filtering in Netflix uses three different data sources in order to generate rec-


Figure 3.5: The General Interface of Netflix

ommendations. In the Table 3.1, the sources and their usage strategies are described. Simply, ranking range is 1 to 5 and it is available to users to comment on the movies.

For generating the recommendations, Netflix firstly searches for similar users that have the same movie history. Considering these similar user information, the collaborative recommendation engine tries to identify the movies that user may like. Netflix claims that %75 of the recommendations is accurate and Root Mean Square Error value is evaluated as 0.8743.

Table 3.1: The Sources of Netflix Recommendation Engine

| Source | Description |
|---|---|
| Movie Data Set | The categorized movie data set based on the movie item content. |
| Queues | The user queues that include the information rented and commented movies. |
| Netflix Ratings | The combined ratings that Netflix predicts with the usage of its recommendation algorithm. |

## 3.5 General Discussion

Last.fm is one of the inspiring projects for this thesis study considering the embedded social circle analysis in its recommendation algorithm. The Last.fm architecture is created as a social network site in which you can create your own friend lists and social groups. Although the information of the users' social networks are not directly used in the recommendation algorithm, the idea of relating recommendation and social network in Last.fm is inspiring for this thesis study.

Beside this when the algorithm of Last.fm is analyzed regardless from the domain information, it is accurate to say that it is an important and successful example for the algorithms that has been studied in this thesis study. In the algorithm design process, some of the basic principles of Last.fm are used.

On the other hand, the FOAF project is another inspiration point for this thesis study considering its name and the rooting idea. The questions that 'Is 'Friend of A Friend' really our friend and is it accurate to say that they can be used in suggestion generations?' are the most important questions of this thesis study. In the designed algorithms, it is tried to be answered these questions in the recommendation area by stating statistical facts.

Additionally, the idea of using existing knowledge that is defined on the social networks is one of the other key aspects that are studied in this study. It is tried to use existing meta data of users on Facebook in order to generate suggestions. By this way, it is aimed to collect required user data without creating too much workload on the users.

Facebook is one of the most important key aspects of this study. Similarly with the Facebook friend recommendation engine, in this study also it is aimed to use the existing mutual connection graph of the users to generate suggestions. The social circles of the users are analyzed with the aim of using the existing user data in the recommendation algorithms.

In addition to this, one of the stated future works of this study is stated as using connection way as an additional input for the recommendation generation process. At this point, Facebook Friend recommender engine is the inspiration point for this idea. The usage of meeting way information in the recommendation algorithm has been firstly used in Facebook itself.

Since a movie recommendation engine is studied through this thesis, Netflix is the most important example. Considering the domain similarity, Netflix supplies a general overview of the correct actions that should be considered while generating suggestions in the movie domain. Although the size of the application that has been studied through this study much smaller than the Netflix environment, the architectural design is inspiring for this thesis study.

Additionally, considering the friend list opportunity in the Netflix social network, it is accurate to say that, it uses social circle information of the users. Although the social circle is not analyzed as a single profile, the idea of using the friends preferences is help this study at the architectural design.

# CHAPTER 4

# METHODOLOGY

## 4.1 Overview

### 4.1.1 Motivation

The success of a recommender system depends on the accuracy of the retrieved preference information from the users. The performance increases as the knowledge extends. The deductions that the recommender algorithms do, are strengthen with the know-how of the likes and dislikes of the users. Although the more application asks to users about their preferences, the more accurate predictions are produced; a limitation should be performed on the acquaintance phase in order to not to annoy them.

The methodologies that allow accessing existing data are searched with the aim of gaining enough information about users without asking many questions to them. As a reachable and confidential way of accessing data, the social networks are stated as the new source of the information.

The main reason of the usage of the social networks in the recommendation area is the availability of the extensive accurate data on these platforms. It is clear to say that with the information that is supplied by the answers of specific questions, it has never been possible to create an accurate profile of the users that mirror the characteristic of their personalities [27]. Besides the possibility of the existence of the misleading answers to the questions, the fixed structure of the questions also limits the information that can be retrieved. By using the existing information on the social networks, the chance of accessing the up to date data about the user profiles is available.

In addition to the stated advantages, the social networks allow to access the social circle profiles of the users by indicating the relation reasons. This reasoning mechanism supplies recommendation algorithms with the clues about the closeness degree between the users and their social circles. On the other hand, physiologically it is accurate to say that family and friends are the most important effects that shape out our personalities. Social networks allow us to access this innovative data without asking or predicting about it.

In this research, it has been studied to recommend movies based on the personal and social circle information of the users that have been supplied through a commonly used Social Network, Facebook. Only Collaborative Filtering approach has been applied in the recommendation generation phase, in order to emphasize the importance of the social environment on people's preferences. The Content-Based Filtering approach has been not considered with the aim of providing more pure and accurate evaluation.

As the social network, it has been decided to use Facebook because of the popularity and reliability of the application. Considering the advantage of accessing personal and social circle information, this platform is stated as the best environment that the algorithms can be applied. Besides, with the motivation of Facebook to its developers, the integration of the application to the platform has been done in an unproblematic and organized way. The benefit of Facebook has been used for reaching more people and extending the test user data set of the application.

### 4.1.2 Recommendation Strategy

Collaborative Filtering methodology emphasizes the importance of the user profiles' similarities on the recommendation generation. In this study, the user profile similarity idea is extended by considering the users' social environments and their families. The user similarity approach on the collaborative filtering is improved to the social environment similarity with the creation of virtual profiles of their social circles. It is aimed to create representations of the social circles of the users considering their movie tastes. With the movie questions regarding to the social circle of the participants a virtual preference set is built for the social circles in the same way that is done for the user profiles.

The approach is rooted from two basic questions and the motivation of the study is determined as answering them in movie recommendation area through a trustful evaluation methodology. The questions are:

- If the movie preferences of two persons are similar to each other, is it accurate to say that the common preferences of their social environments are also similar? Considering the fact that suggestions can be generated from the social circles of the users, is it possible to use the common preferences of these similar social circles for suggestions? In other words, if two persons are like to each other, is that also mean that their social environments are also similar at an acceptable level?

- If the common movie preferences of the social groups of two different people are similar to each other regardless from their personal similarity, is it accurate to say that their movie preferences are also similar? Although these two persons, that have similar social circles, are not similar to each other, is it possible to use preferences of one of them for suggestions to the other one? In other words, if the social environments of two persons are like to each other, is that also mean that they are similar at an acceptable level?

Through this study, the social circle effects on the user preferences are identified by evaluating the result of the corresponding questions. With the data retrieved from users through a movie recommendation application on Facebook, the preference profiles of the users and their social circles are identified.

Considering the large friend sets of Facebook users, with the aim of increasing virtual social profile success, instead of the usage of the entire friend list, a elimination strategy is applied to the social circles. It is implicitly asked to the users for selecting their own close related friends whose movie critics are important and remarkable. The inclusion of these friends is performed by the users themselves by referring the owner of their memorized movie critics that are included in their friend lists.

## 4.2 Domain Description

### 4.2.1 Application Description

The movie recommendation application, ’Suggest Me a Movie’ [49] is published on Facebook platform with the intention of gathering data from the users. Creating a pure and simple interface is identified as the principle of the application. Very flexible question and commenting mechanism is served to the users in order to not to annoy them by struggling for the data retrieval process.

#### 4.2.1.1 Application Profile

The technical information about the recommendation strategy is shared with the curious users on the profile page of the application. In a very clear and brief way, recommendation idea behind the application is stated and the aim of the study is shared with the users. Review and discussion sections are also included in the profile page with the aim of creating more communicative and interactive platform.

A screen-shot of the application profile page one Facebook is given Figure in 4.1.

#### 4.2.1.2 General Instructions Section

The application starts with general instructions for the users who cannot find the time for checking the application profile page. ’General Instructions’ supplies a brief explanation of the aim of the application and clearly states the required number of questions that should be commented in order to guarantee the generation of accurate recommendations. On every stage of the application, friendly language is used for creating more sincere environment.

#### 4.2.1.3 Preferences Section

For achieving user friendliness, a quick version of data retrieval process is served for annihilating the risk of annoying the users. By ’Preferences’ Section, users have the chance of

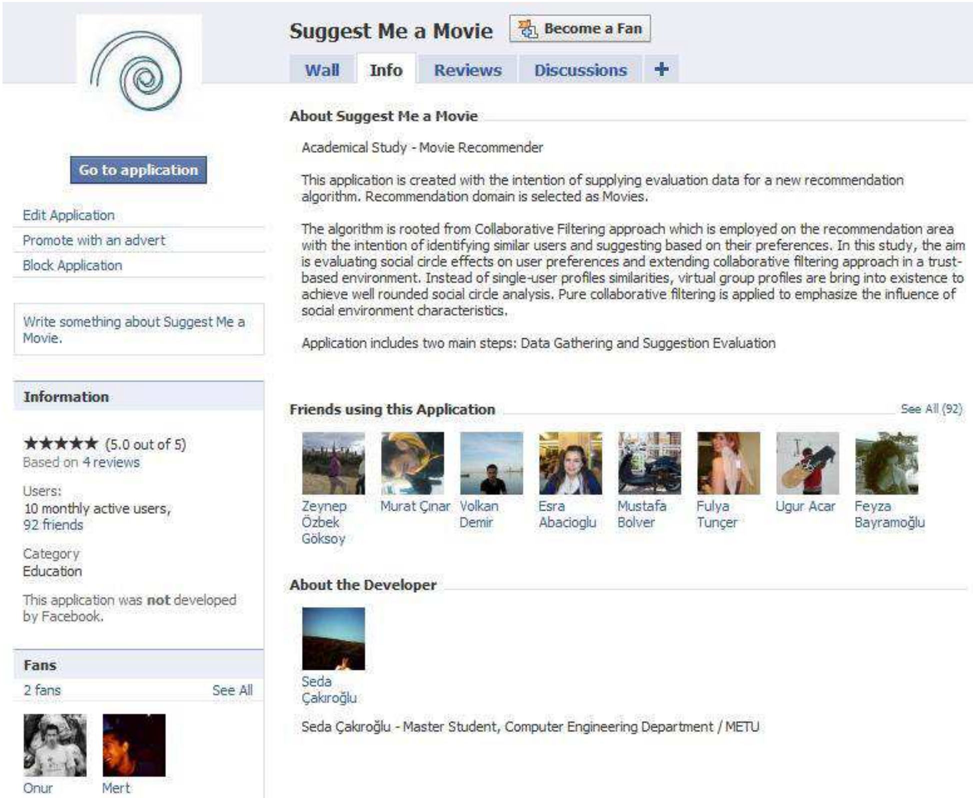


Figure 4.1: The Suggest Me a Movie Application Profile

reviewing ten movies at the same time with the opportunity of commenting only the ones that they prefer.

It is asked to commenting on the movies based on their personal experiences and the movie critics that they hear from their environments. Also in order to supply information about the movies, IMDB links are integrated in the application and done accessible by double clinking on the movie names.

A screen-shot of 'Preferences' section is given in Figure 4.2.

#### 4.2.1.4 Random Movies Section

As the main source of the data retrieval process, a detailed movie question section is included in the application. In the 'Random Movies' section, it is asked to comment on the randomly selected movies based on the preferences of the users and the critics that they have heard from

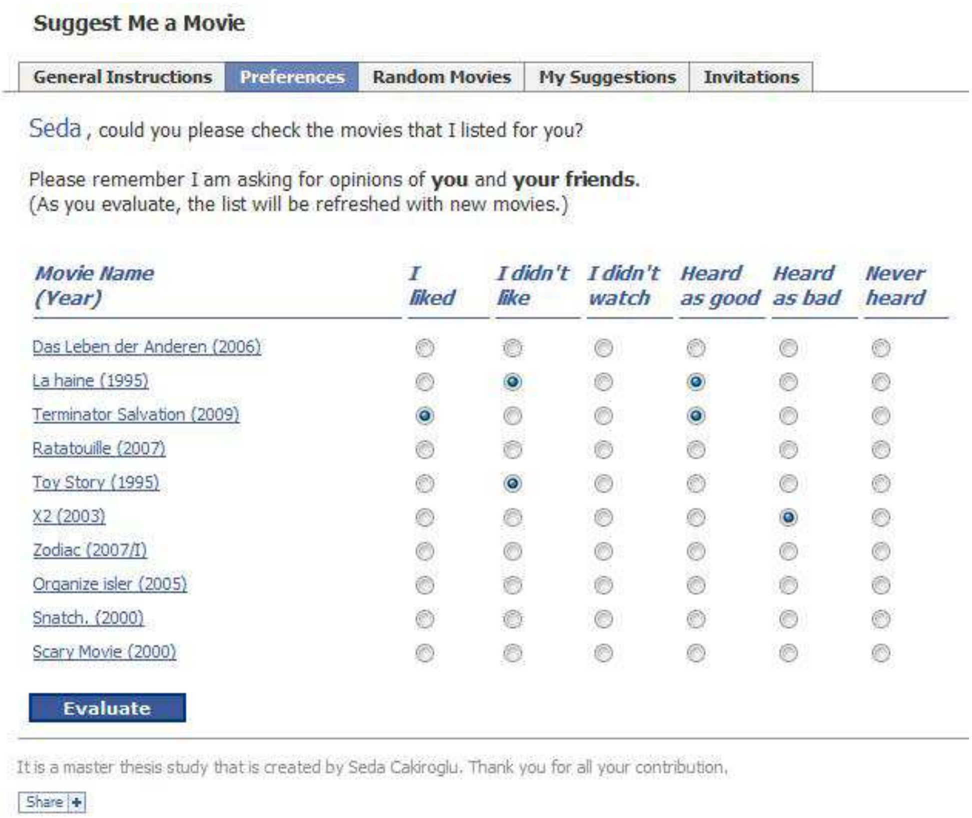


Figure 4.2: Preferences Section

their social environments.

For commenting on the answers that are related with users' social circles, personal referring opportunity is integrated in the application by using the default Facebook functionalities. By this opportunity, it is allowed users to refer their friends personally whose movie critics are important and remarkable for them.

A screen-shot of 'Random Movies' section is given in Figure 4.3.

#### 4.2.1.5 My Suggestions Section

Recommendations that are generated based on the user and social circle profiles are presented to the users in the 'My Suggestions' section of the application. Near the movie information, the generation algorithms of the suggestions are shared with the users in order to supply a

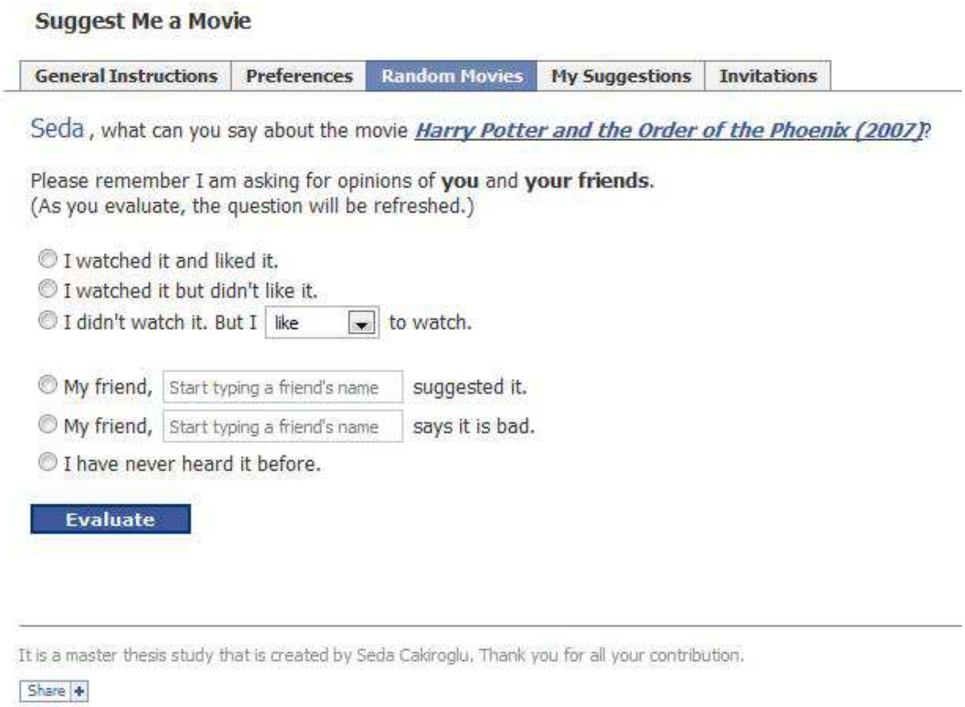


Figure 4.3: Random Movies Section

more clear explanation of the reasons of the suggestions.

Moreover, an evaluation section is added at the end of this part with the aim of understanding how the users fell about the suggestions. In addition to the statistical evaluation data that is evaluated considering the predicted rates and the actual rates of the users, it is assumed that the idea of giving the chance of criticizing the application generates more accurate evaluation of the recommendations.

A screen-shot of the 'My Suggestions' section is given in Figure 4.4.

#### 4.2.1.6 Invitations Section

Additionally, the 'Invitation' section is formed with the aim of increasing the test user data set of the application. By this page, the opportunity of sharing 'Suggest Me a Movie' application with the friends of the participants through Email or Facebook messages is served. Moreover, Facebook share button is placed at the bottom of the each page for allowing users to publish

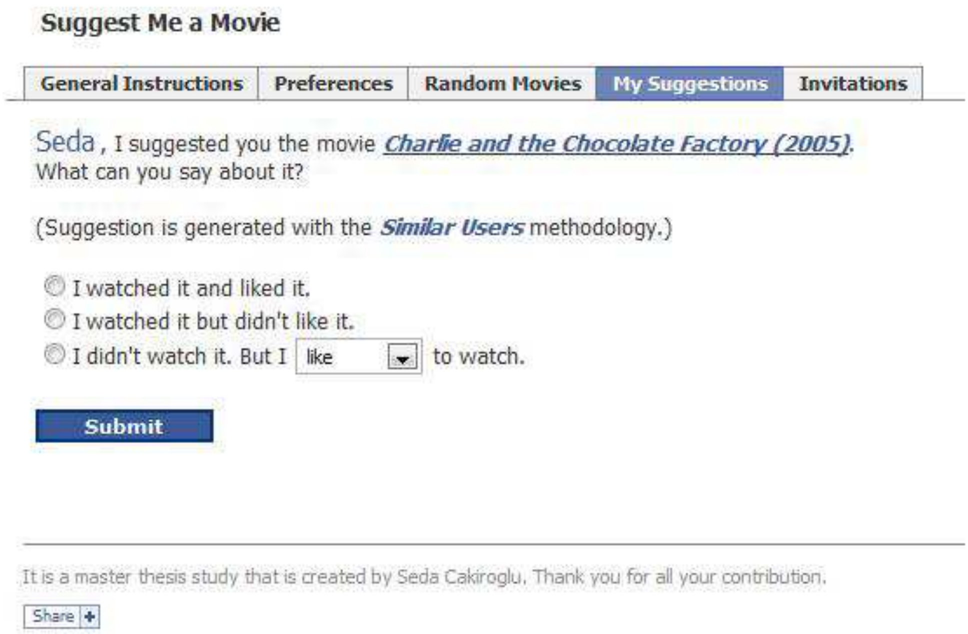


Figure 4.4: My Suggestions Section

the application information on their profiles with their personal comments.

The screen-shot of the 'Invitations' section is given in Figure 4.5.

### 4.2.2 User Profile

Suggest Me a Movie application is published on Facebook with the intention of accessing test users. After publishing the application, it is identified that 119 users accessed and checked it. Analysis shows that only 52 of them have used and supplied a meaningful data for the further evaluation. These 52 participants are selected as the test group for the evaluation phase of the study.

Based on the personal information about the test users which is supplied by Facebook, the user profile of the application is analyzed (Figure 4.6). It is seen that the age range of the test users is 19-45 where the mean age value is 27. Moreover, it is identified that all the users are highly educated, including university, master and doctoral degrees.

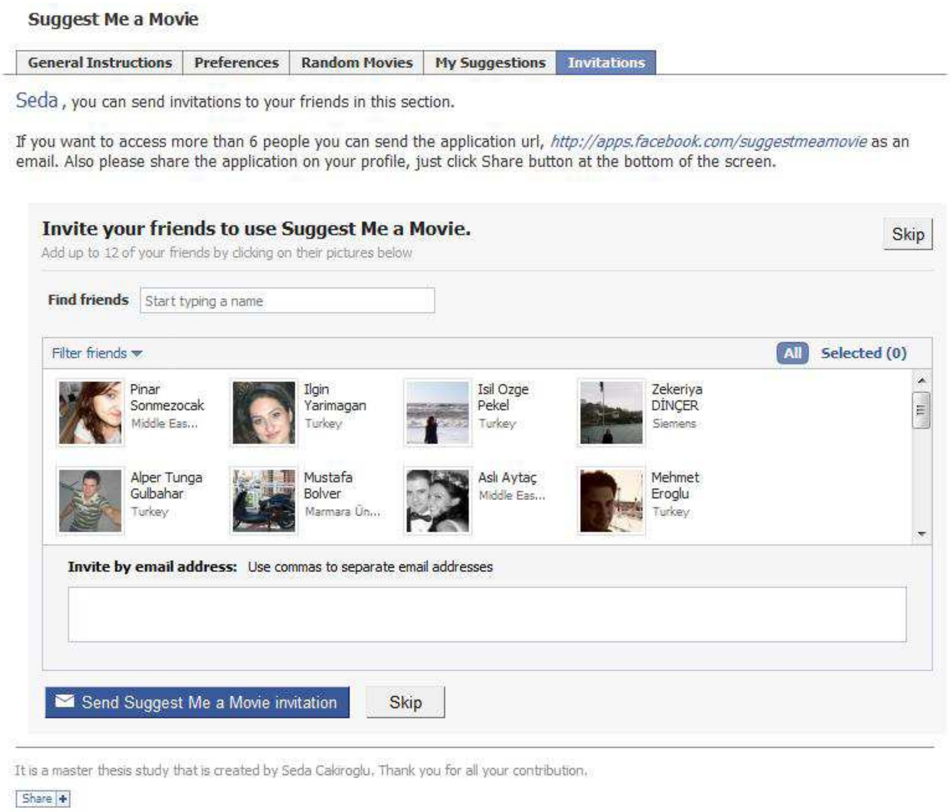


Figure 4.5: Invitations Section

### 4.2.3 Item Profile

Statistically, it is decided to include 310 movies in the data set in order to guarantee useful intersection on the commented movie data for each user. The data is retrieved from IMDB based on year, vote and nationality information. Considering the approaches of the algorithms and the importance of the social circles of users, an accurate interval is tried to be identified that captures the social circle development age of the test users.

Based on the profile mean age value 27, 250 items of the movie data set is selected from the year period 2000-2009. The rest of the items are selected from year range 1990-1999 by the intention of identifying more sophisticated users.

The diagram shows the distributions of the movie items based on their publish year information (Figure 4.7).

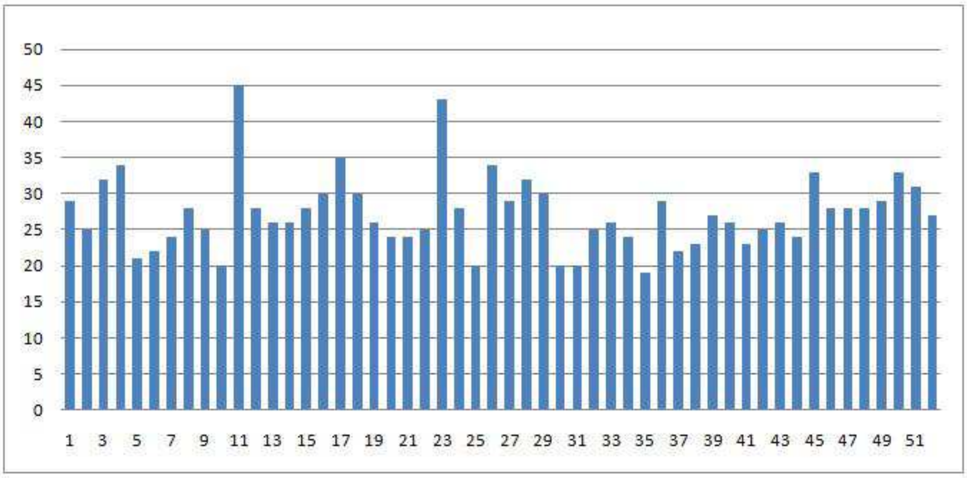


Figure 4.6: Age Distribution of Users

The most popular best and worst movies of the identified year ranges are selected for the movie item data set. High popularity items are included in order to increase the possibility of recognition of the items by the users. A small set of unpopular movie items are also taken part in the item set in order to recognize more sophisticated users.

### 4.2.4 Data Acquisition

The 'Suggest Me a Movie' application is created on Facebook with a simple and pure interface with a view to create widespread data gathering source. It is asked to the users to comment on the movie related questions based on their personal experiences and their memorized movie critics from their social environments.

For all the movie items, it is requested to comment on the available options. However, it is decided to not to include ranking option for the available comments. The reason of this decision can be stated as retrieving more concrete answers from the users and decreasing the diversity between the users' ranking strategies. With the aim of creating personal profiles and virtual social circle profiles, two different comment sets are served to the users; the personal specific comments and the social circle specific comments.

Personal specific comments include; 'I watched it and liked it', 'I watched it but didn't like it', 'I didn't watch it but I like/don't like to watch' (Figure 4.8).


Figure 4.7: Item Set

The social circle specific comments include; 'My friend, suggested it', 'My friend, says it is bad', 'I have never heard it before' (Figure 4.9).

For this part personal referring to the owner of the memorized movie critics is allowed with the aim of identifying close related friends. By this way, the friends who have higher importance for the current user considering the movie taste can be identified. The users have the chance of supplying the name of their friends who have declared an idea about the randomly retrieved movie items. By these comments, it is aimed to identify a small set of user's friends whose ideas have stronger impacts on the movie taste of the user.

### 4.2.5 Data Representation

In this study, two main data structures are used for the generation of the recommendations with the studied algorithms; the user profile and the social circle profile. These data structures are created as vectors for every user that authenticates the application. The creation of the

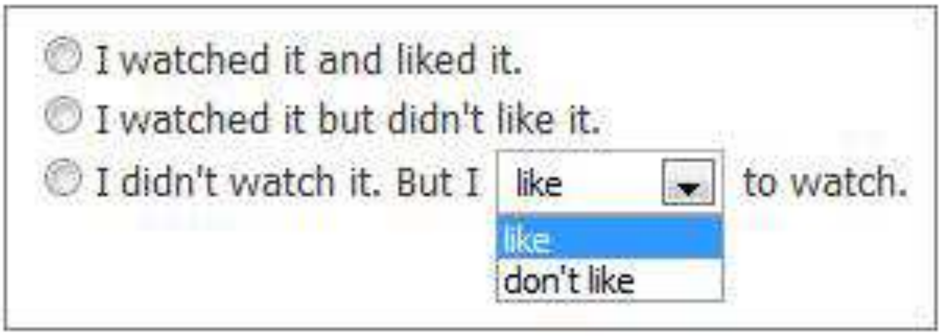


Figure 4.8: Personal Specific Comments

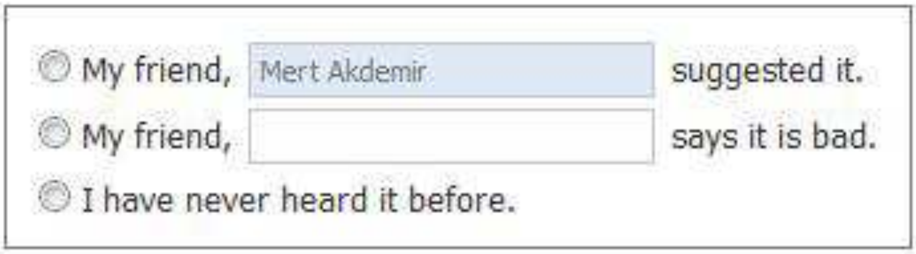


Figure 4.9: Social Circle Specific Comments

user profile vectors and virtual social circle vectors is performed at the permitting phase of the application by which users allow the application to access to all their personal data on their Facebook accounts. With the comments on the presented movies, the enrichment process of the vectors is performed with the predefined ranks of the selected comments.

Table 4.1: Personal Specific Comments and Ranks

| Comments | Ranks |
|---|---|
| I watched it and I liked it. | 5 |
| I watched it and I didn't like it. | -5 |
| I didn't watch it. But like to watch it. | 3 |
| I didn't watch it. But don't like to watch it. | -3 |

The enrichment of user profile is performed based on the selections on the personal related comments. For every comment, the vector is enriched by the stated rank (Table 4.1). When user selects one of the comments from the Personal Specific Comment list, his/her User Profile vector is updated with the corresponding data.

Table 4.2: Social Circle Specific Comments and Ranks

| Comments | Ranks |
|---|---|
| My friend, XXX suggested it. | 5 |
| My friend, YYY says it is bad. | -5 |
| I have never heard it before. | 0 |

Enrichment of the social circle profile is performed based on the selections on the social environment specific comments. For every comment, the vector is enriched by the stated rank (Table 4.2). When user selects one of the comments from the Social Specific Comment list, his/her Social Circle Profile vector is updated with the corresponding data.

In addition to the user and social circle profiles, personal referring is also recorded in order to identify the close related friends of the users. For every user, a relation vector is built that includes the data of the referred users' personal profiles and social circle profiles. For the case that referred user is not an active user of the application, it is decided to not to include this information. They are excluded from the relation vector and that kind of feedbacks are discarded.

## 4.3 Data Storage

The user specific data that is retrieved through Suggest Me a Movie application is stored on a Mysql database in an organized structure for allowing accessing the information quickly and easy way in the recommendation generation phase. In the low level, five data structures are created on the database, which are Users, Groups, Items, Relations and Predictions. Each of the data structure is represented as relational tables on the database level.

Figure 4.10 shows the relation between data objects.

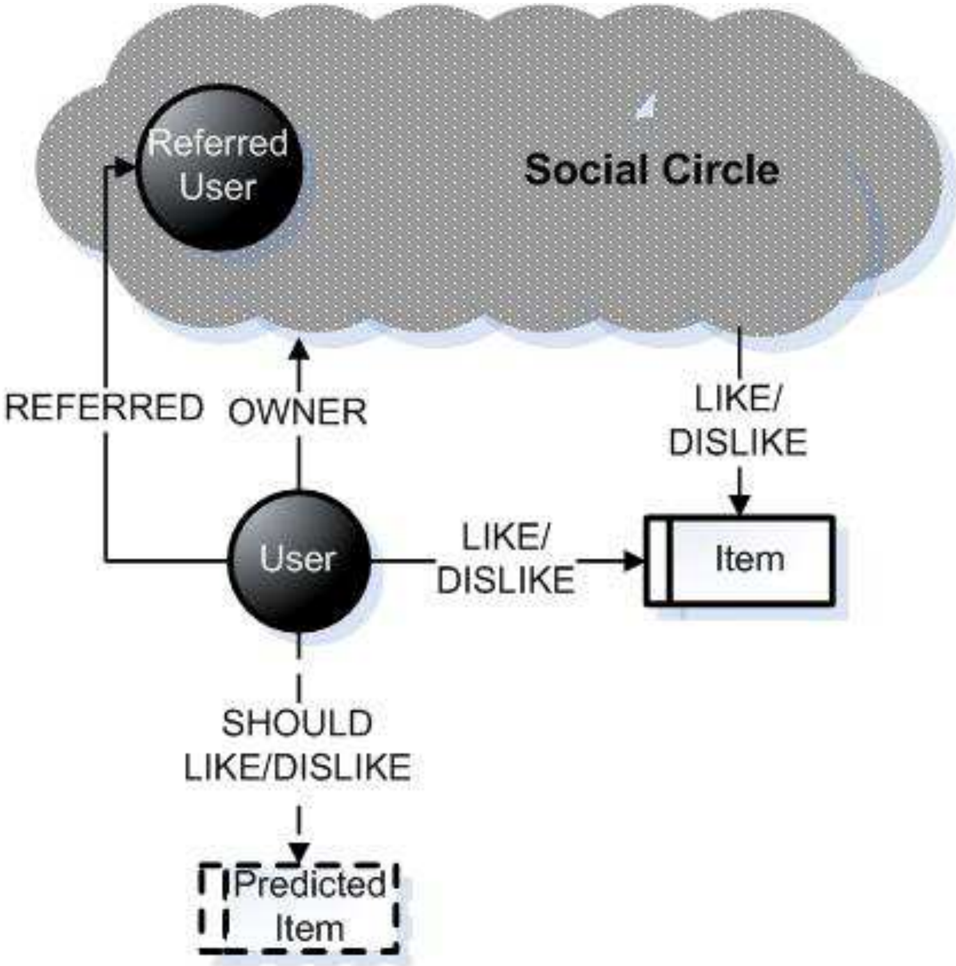


Figure 4.10: Relations between Data Objects

### 4.3.1 Users

User table stores the personal and application related general information of the users. Near the saved 'FirstName' and 'LastName' information of the users, it saves the 'startid' and 'currentid' values in order to infer the movie item ids that user accessed until now.

In the random movies and preferences section of the application, instead of serving the movie items in a totally random or ordered way, it is decided to serve users a small ordered set of the movies. Assuming that every user comments on approximately 40 movies, the described structure is used in order to increase the common item amount in the commented movie sets of the users.

### 4.3.2 Groups

Groups table represents social circles of the application users. It saves ’groupid’ as the identifier of the virtual social circle profile and foreign key ’userid’ which indicates to the owner users of the social circles.

### 4.3.3 Relations

Relations table saves the personal referring information that is supplied in the social circle related comments. With the user names that users specified as the owner of their movie critics, are saved in this table in order to access the closely related friends’ information of the users. The ’Userid’ and ’Facebookid’ fields are available on this table.

### 4.3.4 Items

Items table stores the movie information that is served to the users through the application. Near ’itemid’ data, the ’name’ and the IMDB ’URL’ of the movie are saved on this table.

### 4.3.5 User Ratings and Group Ratings

These tables store the ranks that users supply as the answers of the randomly served movie questions. User Ratings table saves the ranks of the user specific comments whereas Group Ratings saves the ranks of the social circle specific comments. Each of the tables is stored as items versus user/ group vectors.

### 4.3.6 Predicted Similar User Rates and Similar Group Rates

These tables store the predicting ratings for the users for the unrated movie items. Since two algorithms are applied through this study as ’Recommendation Generation Based on The Social Circles of Similar Users’ and ’Recommendation Generation Based on The Similar Social Circles’, the results are saved in two different structures. ’Userid’, ’itemid’ and ’predictedrate’ fields are available on these tables. Also an additional field, ’realrate’ is added in order to store the actual ranks that users supply to the recommended items.

### 4.4 Recommendation Algorithms

The goal of a recommender system is to help users to find the desired information objects [50]. The collaborative filtering methodology emphasizes the importance of personal profiles' similarities on the recommendation generations [51]. With the aim of increasing the recommendation success, the tastes and preferences of the similar persons are used for the suggestion generations. In this study, it is aimed to extend this idea by considering the social circles of the peoples.

Researches on psychological area show that the families and friends are the most important impacts that shape out our personalities and preferences. Since our tastes are rooted from our environment, it is accurate to say that our social environments are the presenters of our basic preferences because of the common social and cultural environments. Considering this physiological referring, through this study the methodology of recommending with the usage of social circles information of the users is studied in two different approaches and the success rate of these approaches are evaluated.

Moreover, considering the application platform, Facebook, it is clear to say that elimination should be performed on the friend lists of the users in order to increase the trust of the application methodology. Instead of the usage of the entire friend list that also includes unrelated persons; a filtering approach is applied by serving social circle specific comments to the users in addition to the personal specific comments. By this strategy it is aimed to be performed the elimination by the users themselves and increasing the trust of the applied algorithms.

#### 4.4.1 Recommendation Generation Based on The Social Circles of Similar Users

Recommendation systems that are available on the web are using very similar algorithms just including small modifications. However, since this is the case, users are bored to see the same suggestions in every recommendation engine. Considering this, in this study, the question that 'friend of a friend is it also our friend?' is asked in order to create more innovative suggestions.

Currently friend of a friend idea is used in many areas as the social networks grow. The idea is applied on the friend recommendation engines and the success is improved in the trust based environments. In this study it is tried to use this methodology considering the social circle information.

Instead of using friends as a single source of information, a more general perspective is tried to be created by representing all the friends in a single virtual social circle profile. By this way, it is aimed to create a more common profile that presents all the friends of a user. Considering that similar users can be used for generating suggestions, the idea of using the virtual social circle profiles of these similar users for generation of the recommendations is investigated.

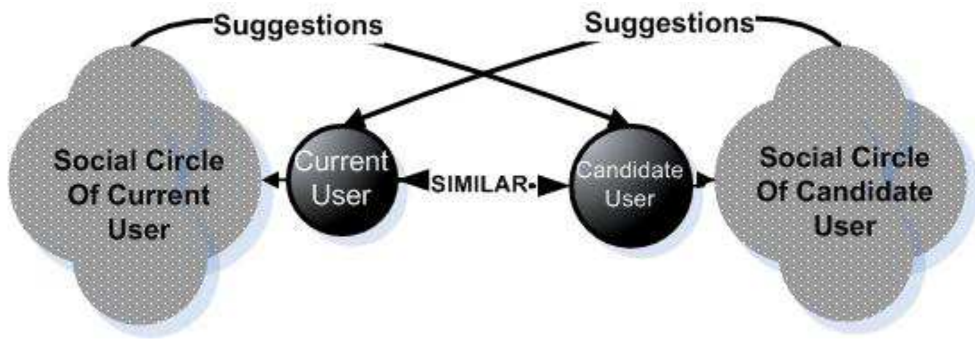


Figure 4.11: Idea Behind The Similar Users Algorithm

In the 'Suggestion Generation Based on The Social Circles of Similar Users' methodology (Figure 4.11), the suggestions are tried to be generated considering the movie taste of the social circles of the similar users. After the identification of the similar users, although the preference data of these users are available, only the social circles of them are evaluated for creating innovative and extensive recommendations. Regardless from their personal profiles, the virtual social circle profiles that are formed with the data that the users supply through the social circle specific comments are used with the intention of investigating the importance of the social environments impact on our personal preferences.

Additionally, in order to emphasize the owners of the user's memorized movie critics, the virtual social circle profiles of personally referred friends are also included in the social circles list if an active application account is available for them. By this way, it is aimed to increase the importance of the social circles of that referred friends. Since it is clear to say that these referred friends are the most important factors on the movie taste of the user, the social circle preferences of these referred friends; should be also emphasized in the algorithm.

### 4.4.2 Recommendation Generation Based on Similar Social Circles

The social environments are the presenters of our personalities and preferences based on the social and cultural reasons.

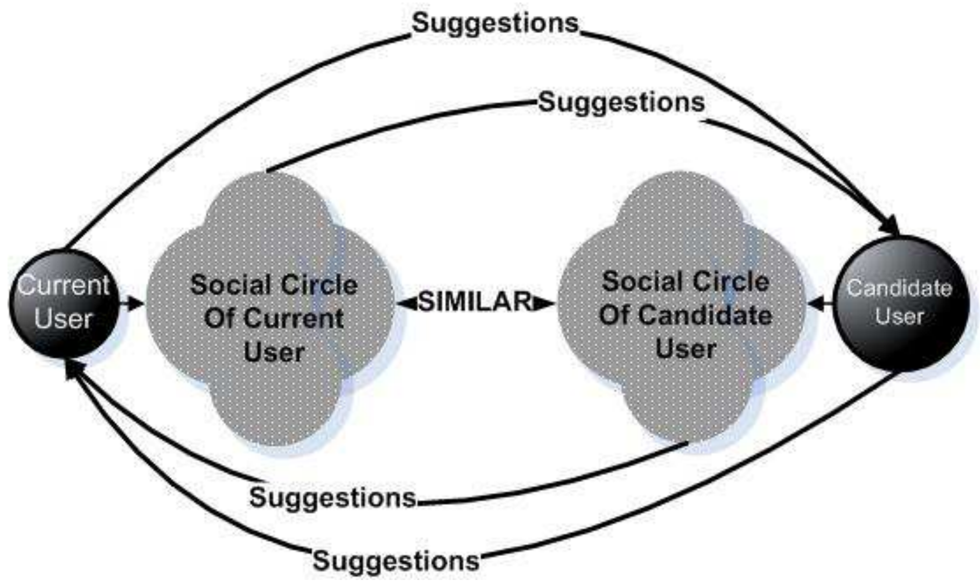


Figure 4.12: Idea Behind The Similar Social Circles Algorithm

In 'Recommendation Generation Based on The Similar Social Circles' methodology (Figure 4.12), the similarity degree between the social circle profiles of two users is investigated instead of the user similarity. The idea of user similarity on the collaborative filtering approach is extended by the consideration of the social circles similarity.

The similar preferences are tried to be identified by only including common tastes of the social environments of the users. Regardless from the personal information that users supply through the Suggest Me a Movie application, only the social environment specific data is used in order to form the virtual social circle profiles of the users. The recommendations are generated considering basic similarities on these virtual social circle profiles.

Additionally, the owner users of the similar social circles are included in the recommendation generation process, by considering that these users are also members of their social circles. After the identification of the similar social circle profiles, the profile information is enriched with the owner users of the selected similar social circles.

Moreover, in order to emphasize the effect of social circles of personally referred friends, regardless from their social circles' similarity to the users' social circles, the virtual social circle profile of them are also added to the similar social circle list. The importance of these social circles is increased in the evaluation process with the aim of increasing trust of the recommendation algorithm by using the data that is served directly by the users.

# CHAPTER 5

# SYSTEM ARCHITECTURE

## 5.1 Personal Profile and Virtual Social Circle Profile Identification

In the cold start, the movies that are served randomly in the Preferences and Random Movies section of the application is selected in an organized way instead of using completely random or ordered methodology. For every user a random start id is selected and the movies that are served are retrieved in an ordered way, starting from that randomly selected start point. This strategy is used in Suggest Me a Movie application in order to generate a better distribution on the rated movie data set for the evaluation process of the algorithms.

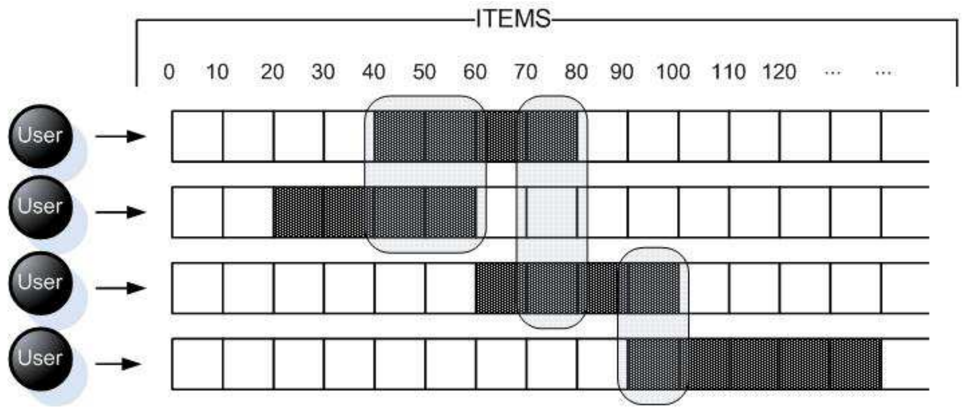


Figure 5.1: Distribution of Randomly Retrieved Movie Items

Figure 5.1 illustrates the random movie item strategy of Suggest Me a Movie application. As stated in the diagram, firstly a start point in the movie item data set is selected for every user. After that point, the movie items are served in an ordered way to the users. By this way it is aimed to increase the amount of the commonly rated movie items between the users.

### 5.1.1 Enrichment of The Personal and Social Circle Profile

For the shape out process of the personal profiles and virtual social circle profiles, two different comment sets are served to the users; personal specific comments and social circle specific comments. Without noticing the user that collection of the social circle preferences are performed, only general comments that indicate the memorized critics about the items are asked to the users.

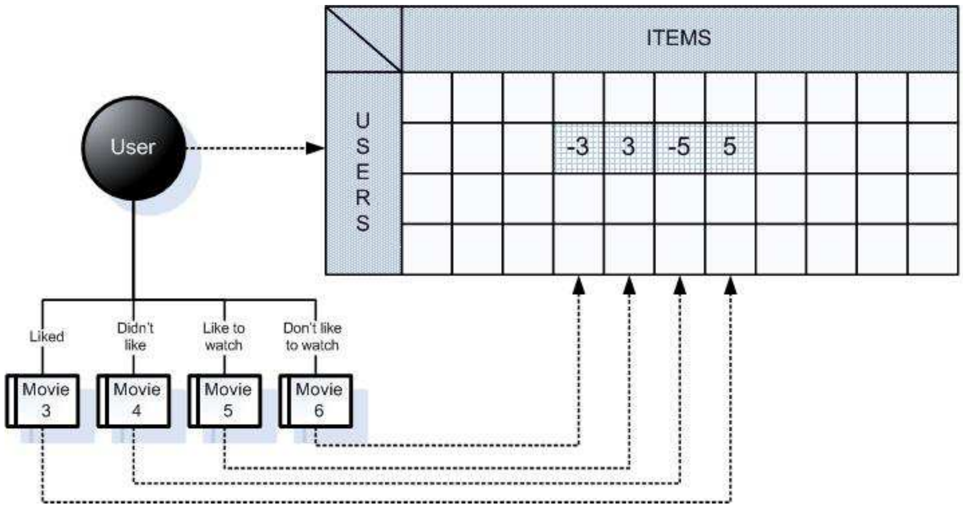


Figure 5.2: User Profile Update Based on Feedbacks

With the selection of the comments in the user specific comments set, the personal profile vector is updated with the corresponding data (Figure 5.2).

On the other hand, with the selection of the comments in the social circle specific comments set, the group profile vector is updated with the corresponding data (Figure 5.3).

### 5.1.2 Identification of Referred Users

The personal referring functionality is applied with the aim of filtering the friend list in the view point of the users according to the impact on their memorized movie critics. The relation

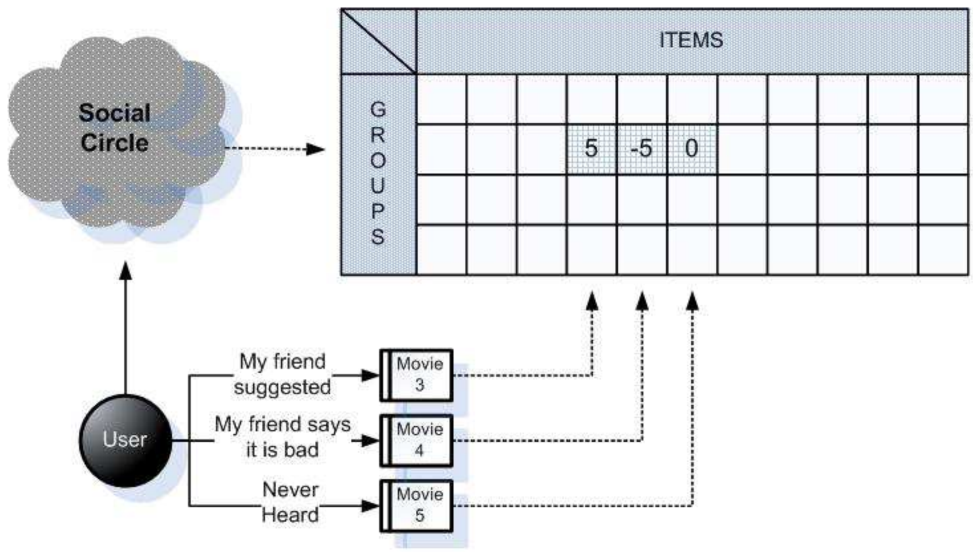


Figure 5.3: Social Circle Profile Update Based on Feedbacks

identification process in the application is served to the users in order to access the information of the friends who are more important than the others considering their movie tastes.

With this intention, the users whose ideas about the movies are cited are recorded to the database in order to use this personal specific information in the algorithm evaluation.

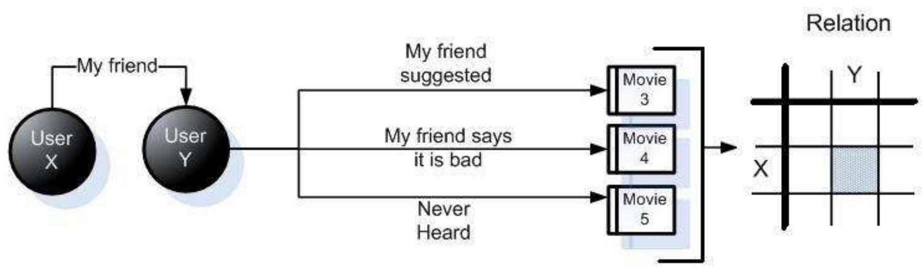


Figure 5.4: Relation Information Update Based on Feedbacks

With every personal referring the relation table is updated with the corresponding data (Figure 5.4).

## 5.2 Recommendation Algorithms

### 5.2.1 Recommendation Agent - Recommending Based on The Social Circles of Similar Users

Considering the comments that are supplied in the cold start phase of the application, for every application user, a similar user list is identified in which the users who have highest similarity rank, are included. After the identification of that similar user list, the social groups of the users who are included in the list are selected as the evaluation group list. Also with the inclusion of the social circles of the referred users, the evaluation group list is finalized for the prediction phase.

In the prediction phase, predicted rank for every unrated item is evaluated for every application user.

Here is a representation of idea behind the Social Circles of Similar Users algorithm. In the diagram the bobble that is named as current user symbolizes the user for who suggestions will be generated. On the other hand, the bobble Candidate User symbolizes any random user that share the same movie taste with the current user as stated in the diagram. With the identification of the similarity between the current user and the candidate user, the social circle profile of the candidate user is used for the generation of the recommendations. For the current user, the suggestions are generated based on the data of the social circle of the candidate user instead of the personal profile data of the candidate user itself directly (Figure 5.5).

The description of the algorithm is shared step by step in Table 5.1. In the sections below, detailed descriptions of the algorithm's steps are included.

#### 5.2.1.1 Similarity Rank Calculation for Users

For the evaluation of the similarity ranks of the user profiles, Pearson Correlation Coefficient Factor (PCC) is calculated for every user that authenticates to the application separately. PCC factor is one of the most commonly used similarity decision equation in the recommendation

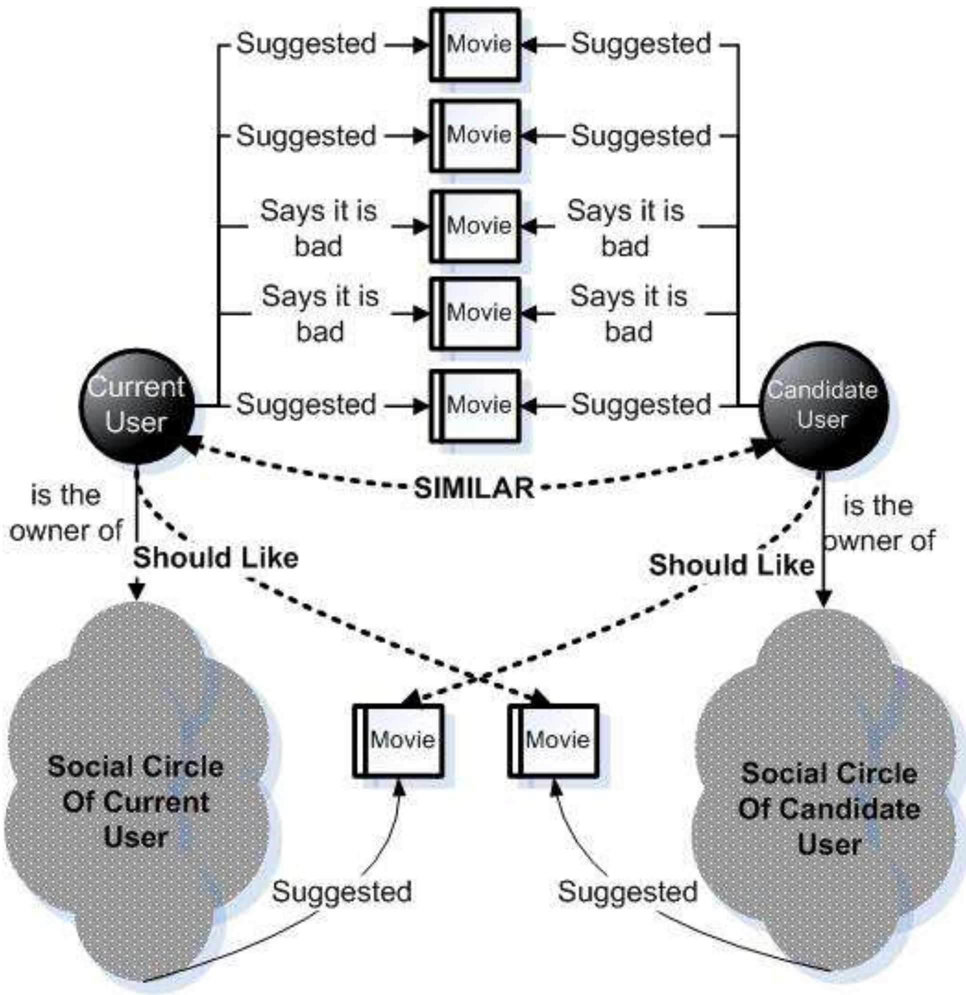


Figure 5.5: Recommendation Based on Social Circles of Similar Users

systems which supplies concrete data for the suggestion generations. With the usage of this factor an accurate identification of the similar users are tried to be performed.

Pearson Coefficient Correlation factor is a statistical correlation factor that measures the strength of the linear dependency between two vectors with the consideration of the sensitivity of the data distribution [53]. The idea is firstly presented by Francis Galton in 1880 and afterwards, Karl Pearson enhanced the coefficient. The range of PCC factor is +1 to -1, where +1/-1 indicates the perfect correlation, which means highest dependency between the vectors. As a property of the equation, symmetry can be declared which means the result of the PCC between X and Y and the PCC between Y and X are the same.

Table 5.1: Algorithm Description of Recommendation Based on Social Circles of Similar Users

| ID | Description |
|---|---|
| 1 | Retrieve user profile vector of the current user. |
| 2 | Calculate PCC value between current user profile vector and all user profile vectors in the application. |
| 3 | Order the PCC values, select the highest %20 PCC value and add the owner users of corresponding vector to the similar users array. |
| 4 | Add all the users that are referred in the application to the similar users array. |
| 5 | Find the social circle vectors of the users in the similar user array. |
| 6 | Predict a rate for the included items in the selected social circle vectors. Ensure that they are not rated by the current user. |
| 7 | Order the items based on their predicted rates. |
| 8 | Start suggesting from the highest rated item. |

Table 5.2: Assumptions of PCC Estimation

| ID | Description |
|---|---|
| 1 | There should be no dependency between vector. |
| 2 | The attributes of the vectors should be well distributed. |
| 3 | The vectors should be in a relation of cause and effect. |
| 4 | The vectors should be linear. |

However in order to use PCC Factor estimation in the similarity decision, the assumptions that are listed in Table 5.2, should be applicable for the data set. Since it is accurate to say that all the assumptions are valid for user and social group data sets of this study, it can be concluded that Pearson Coefficient Correlation Factor is applicable for the similarity decision.

$$PCC_{xy} = \frac{\sum_{h=1}^{n'} (r_{u_x,i_h} - \overline{r_{u_x}}) - (r_{u_y,i_h} - \overline{r_{u_y}})}{\sqrt{\sum_{h=1}^{n'} (r_{u_x,i_h} - \overline{r_{u_x}})^2} \sqrt{\sum_{h=1}^{n'} (r_{u_y,i_h} - \overline{r_{u_y}})^2}} \tag{5.1}$$

Where,

- $u_x$ and $u_y$ is the vectors of x user and y user.
- $PCC_{xy}$ is the similarity rank between x and y.
- $n'$ is the total number of the items that are rated in $u_x$ and/or $u_y$.

- $r_{u_x,i_h}$ is the rating that x gives to item $i_h$.

- $\overline{r_{u_x}}$ is the average rating of x for all the rated items.

The equation (Equation 6.2), is applied on the user profile vectors with the aim of calculating their similarity ranks between every pair of user in the application. Since the PCC factor is estimated between every user separately, the complexity of the equation is considerably high, which is $\Theta(n^2)$.

#### 5.2.1.2 Recommendation Data Formation

The user profiles are processed considering their calculated similarity ranks in order to identify the most similar users. Since there is no chance to perform trial versions on the users of the application for identifying an accurate threshold value, the users with the highest similarity ranks are selected directly.

Considering the user count of application, 52, it is decided to select fifteen users in the similar users list for every user. With this aim, the top fifteen PCC values are selected for every user and the owner user profiles of these PCC values are added into the similar users list.

For the formation of recommendation data set (Figure 5.6), for every user, the social circles of the users who are in the similar users list are selected. Without applying any similarity elimination process on these social circle vectors, all the vectors are added to the recommendation data set. The recommendation data set is created as a set of social circle vectors of the most similar users.

#### 5.2.1.3 Enrichment of Recommendation Data with Referred Users

Considering the referred user data that is gathered through the application usage, it is decide to enrich the data of the recommendation data set based on users' personal indications. As described before, since it is accurate to say that these referred users are important in the view of their movie tastes; these users' opinions should be included in the recommendation data set if valid application accounts are available for them.

Two different approaches are considered at this point; including referred user profile vectors

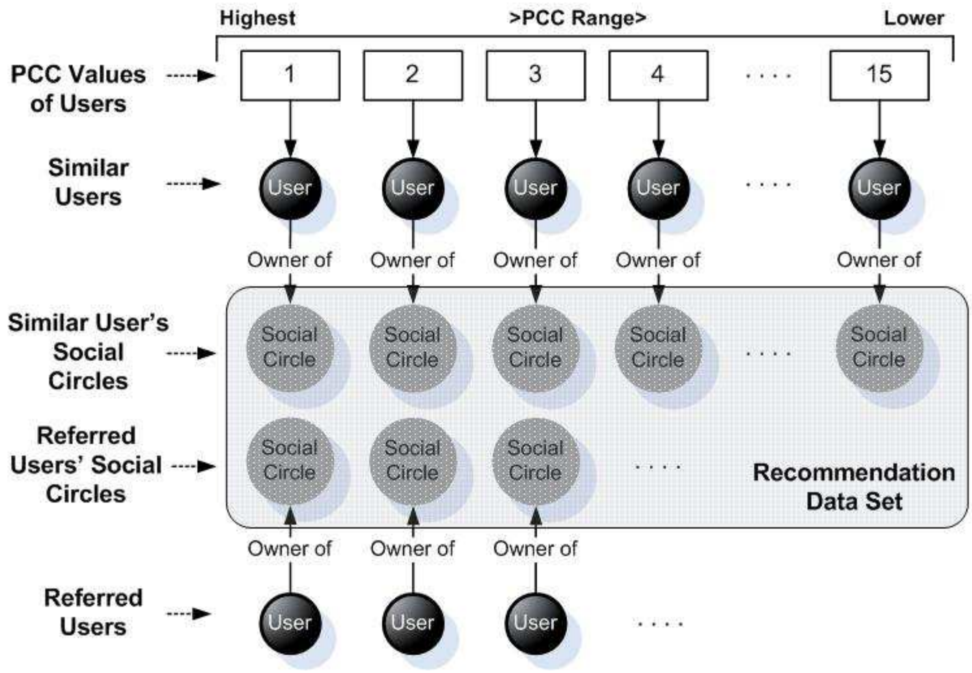


Figure 5.6: Recommendation Data Selection

directly in the recommendation data set or including the social circle profiles of these referred users in the recommendation data set.

Table 5.3: Comparison between Two Stated Approaches

| | Advantages | Disadvantages |
|---|---|---|
| User Profiles | Assurance of Accurate Suggestions | Usage of Already Known Data |
| Social Profiles | Chance of More Innovative Suggestions | Probability of Inaccurate Suggestions |

The table states the advantages and disadvantages of the declared approaches (Table 5.3). Taking into consideration the listed items in table, in the Recommendation Based on The Social Circles of Similar Users algorithm, it is decided to apply the second approach. The social circle vectors of referred users are included in the recommendation data set with the aim of serving more innovative suggestions to the users without the usage of the existing data that is submitted by the user directly. By this approach, the enrichment of the recommendation data set is extended with the addition of the referred users' social circles to the similar users' social circles.

#### 5.2.1.4 Calculation of Predicted Rates

On recommendation data set of the social circles of the similar users algorithm, a prediction algorithm is used in order to calculate a rate for the unrated items. The predicted ratings are calculated separately for every user for every unrated item. The top 8 items that have the higher prediction rates are shared with the users as the recommendations of the application for the stated algorithm.

$$\hat{u}_{ij} = \overline{U_i} + \frac{\sum_{j=1}^{k} PCC_{mi}(u_{mj} - \overline{U_m})}{\sum_{j=1}^{k} |PCC_{mi}|} \tag{5.2}$$

Where

- $m$ is a randomly selected social circle vector that is included in the recommendation data set.
- $\hat{u}_{ij}$ is the predicted rating that user i gives to the item j.
- $u_{mj}$ is the rating that is included in social circle vector m.
- $\overline{U_i}$ is the average of user i's ratings.

In the Equation 5.4, the calculation that is used for the rate prediction is stated. The stated equation aims to predict a rate that the user i is likely to give to the item j. The equation basically uses the similarity rate between the users and the actual supplied ratings of the user. The predictions are calculated linearly with the usage of this equation [54].

#### 5.2.1.5 Implementation Details

The script below, is clearly states the implementation details of recommendation data supply approach. Every part of the implementation is indicated separately in the pseudo code.

In the Figure 5.7, the beginning part of the script is included. The script starts with the evaluation of the PCC values between every user pair. After this estimation, the 15 users that have the highest PCC values are selected for every user.

```
IDENTIFICATION OF SOCIAL CIRLES OFSIMILAR USERS
          (User-Set Uset, Group-Set Gset, Referred-Set Rset,
          User-Ratings Urating, Number-of-Users numb)

START
          // number of users
          N =Length of Uset
          //the number of select similar users
          SEL = 15
          // the estimated PCC value array
          PCC[N + 1] [N + 1] = NULL
          // ordered list of PCC values
          Ordered-List PCC_Ordered[N]
          // similar users array
          Similar-User-List SimilarUsers[N+1][SEL]

          FOR numb, X ϵ Uset
                    // for every user, estimates the PCC Value
                    FOR numb, Y ϵ Uset and Y != X
                              PCC[X][Y] = calculate PCC between Urating[X] and Urating[Y]
                              PCC_Ordered[X] = add user Y and corresponding PCC value
                    ENDFOR
                    // select the top similar 15 users for every user
                    FOR N, i < SEL
                              SimilarUsers[X][i] = owner user of PCC_Ordered[i]
                    ENDFOR
          ENDFOR
```

Figure 5.7: Recommendation Data Set Formation for The Social Circles of Similar Users Algorithm - Part 1

In the Figure 5.8, the second part of the script is included. In this part, the referred users information is retrieved. After that recommendation data set is formed with the social circles of the top similar users and referred users.

### 5.2.2 Recommendation Agent - Recommending Based on Similar Social Circles

In this recommendation algorithm, it is decided to use the virtual social circle profile vectors of the users in order to generate suggestions. Considering the social circles are presenters of the common movie taste of the users based on social and cultural reasons, the similarity decision is performed on the social circle vectors of the users.

For every social circle profile vector in the application, the similarity rank is calculated between every other social circle profile vector. After that calculation, the social circle vectors that have the highest similarity ranks are selected and added to the similar social circles list. This similar social circles list is used as the recommendation data for the further steps. Addi-

```
        // number of referred users for every user
        M = Length of Rset
        // referred users set
        Referred-Social-List ReferredSocialCircles[N][M]
        FOR N, Z ϵ Uset
                // select social circle of referred users
                FOR M, k < M[z]
                        ReferredSocialCircles[Z]= social group of referred user, Rset[k]
                ENDFOR
        ENDFOR

        // evaluation dataset
        Evaluation-Data SocialCircles[N][SEL+M]
        FOR N, Z ϵ Uset
                // include selected social circles of similar users
                FOR SEL, j < SEL
                        SocailCircles[Z][j] = Social Circle of user SimilarUsers[Z][j] from Gset
                // include social circles of referred users
                FOR M,  j<SEL+M[Z]
                        SocialCircles[Z][j]=ReferredSocialCircles[Z][j-SEL]
                ENDFOR
        ENDFOR

END
```

Figure 5.8: Recommendation Data Set Formation for The Social Circles of Similar Users Algorithm - Part 2

tionally the enrichment of the recommendation data set is performed by the inclusion of the owner user profile vectors of the selected similar social circle profile vectors.

In the figure 5.9, the idea behind the algorithm is stated. The bobble, which is named as current user, indicates the user for who the recommendations will be generated. The other bobble, which is named as candidate user, is a random user from the application whose social circle shares the common movie preferences with the social circle of the current user. After the identification of the similarity between the social circles of these users, suggestions are generated based on the information of the social circle of candidate user and the candidate user itself.

#### 5.2.2.1 Similarity Rank Calculation for Social Circles

The similarity rank between the social circle profile vectors is calculated in the same way with the user profile vectors. PCC factor is estimated between the social circle profiles in order to obtain the consistency between similar social circle decision and similar user decision.

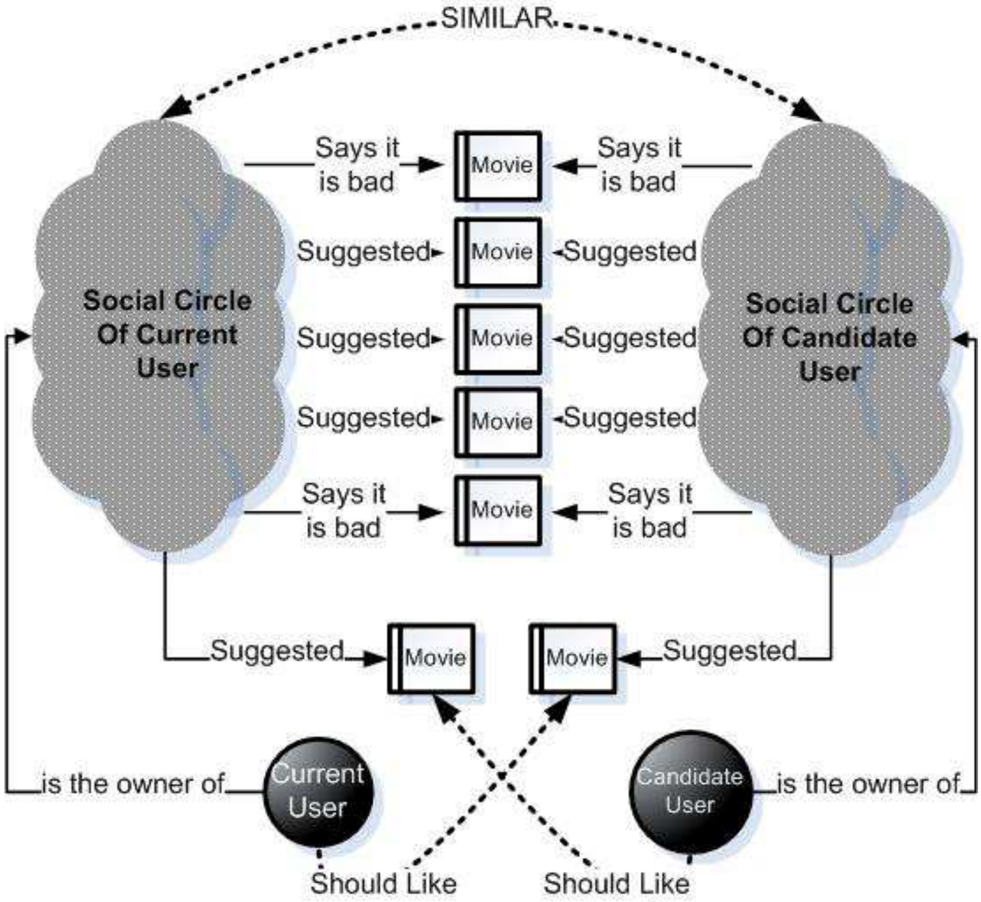


Figure 5.9: Recommendation Based on Similar Social Circles

As one of the other innovative aspect of the studied algorithms, PCC value estimation is used for the calculation of the similarity rank of virtual profiles that represents a social group that includes more than one user. With the consideration of the PCC factor usage assumptions which are dependency, distribution, relation and linearity, it is accurate to say that this approach is applicable for the similarity rate calculation of virtual social circle profiles.

$$PCC_{xy} = \frac{\sum_{h=1}^{n'}(r_{g_x,i_h} - \overline{r_{g_x}}) - (r_{g_y,i_h} - \overline{r_{g_y}})}{\sqrt{\sum_{h=1}^{n'}(r_{g_x,i_h} - \overline{r_{g_x}})^2}\sqrt{\sum_{h=1}^{n'}(r_{g_y,i_h} - \overline{r_{g_y}})^2} \tag{5.3}$$

Where,

- $g_x$ and $g_y$ is the vectors of x social group and y social group.
- $PCC_{xy}$ is the similarity rank between x and y.

- $n'$ is the total number of the items that are rated in $g_x$ and/or $g_y$.
- $r_{g_x,i_h}$ is the rating that x gives to item $i_h$.
- $\overline{r_{g_x}}$ is the average rating of x for all the rated items.

The equation (Equation 5.3), is applied on the social circle profile vectors with the aim of calculating their similarity ranks for every pair of social circle in the application. Since the PCC factor is estimated between every social circle separately, the complexity of the equation is considerably high, $\Theta(n^2)$.

#### 5.2.2.2 Recommendation Data Formation and Enrichment

Consistently with the similar user selection approach, the most similar social circles are selected as the recommendation data for the suggestion generations. For every user, fifteen social circles are selected which have the highest PCC values.

Considering the fact that owner users are also members of their social circles, the enrichment of the recommendation data is performed with the addition of the owner user profile vectors of the selected similar social circle vectors. Regardless from the similarity rate of the owner user profiles, they are added to the recommendation data for extending the data of the selected social circle profiles.

In the Figure 5.10, the idea of formation of recommendation data is declared. As it is seen, the data is a combination of the social circle profile vectors and user profile vectors.

#### 5.2.2.3 Calculation of Predicted Rates

For the similar social circles algorithm, on the recommendation data set, an equation is used in order to calculate a predicting rate for every unrated item. The predicted ratings are calculated separately for every user for every unrated item in the same methodology with the social circles of similar users algorithm. The top 8 items that have higher predicted rates are shared with the users as the recommendations of the application for the stated algorithm.

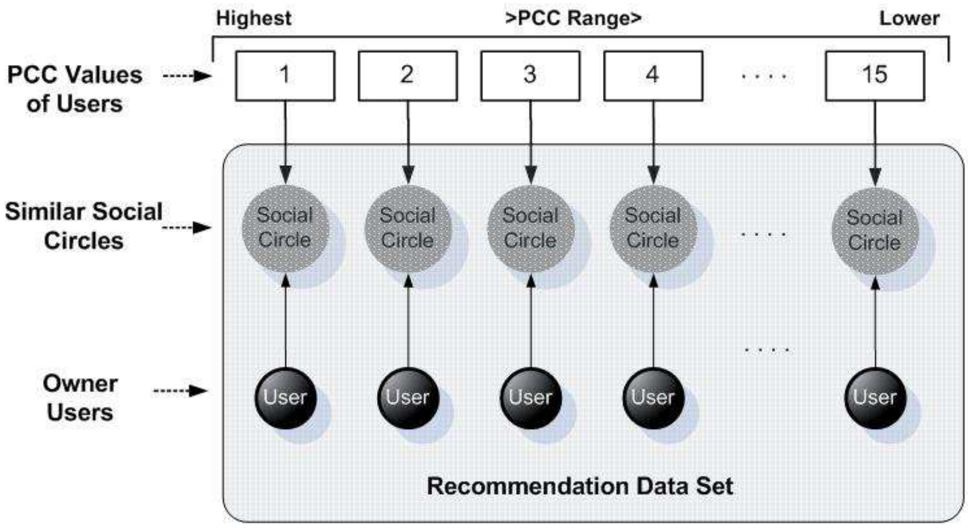


Figure 5.10: Recommendation Data Selection

$$\hat{u}_{ij} = \overline{U_i} + \frac{\sum_{j=1}^{k} PCC_{mi}(u_{mj} - \overline{U_m})}{\sum_{j=1}^{k} |PCC_{mi}|} \tag{5.4}$$

Where

- $m$ is a randomly selected social circle or user vector that is included in the recommendation data set.
- $\hat{u}_{ij}$ is the predicted rating that user i gives to the item j.
- $u_{mj}$ is the rating that is included in vector m.
- $\overline{U_i}$ is the average of user i's ratings.

In the Equation 5.4, the evaluation that is used for predicting the rates is stated. The equation aims to predict a rate that the user i is likely to give to the item j. It uses the similarity rate between the users/groups and the supplied actual ratings of the users. The predictions are calculated linearly with the usage of this equation [54]. In the equation, the social circle vectors and user profile vectors are used in the same way without any discrimination.

## 5.3 Implementation Details

The script below is clearly states the implementation details of recommendation data supply approach. Every part of the implementation is indicated separately in the pseudo code.

```
IDENTIFICATION OF SIMILAR SOCIAL CIRCLES
        (User-Set Uset, Group-Set Gset, User-Ratings Urating)

START

        // number of users
        N = length of Uset
        // the number of selected similar social circles
        SEL = 15
        // the estimated PCC value array
        PCC[N + 1] [N + 1] = NULL
        // ordered list of PCC values
        Ordered-List PCC_Ordered[N]
        // similar social circles array
        Similar-SocialCircle-List SimilarSocialCircles[N+1][SEL]

        FOR numb, X ϵ Gset
                // for every social circle, estimates the PCC value
                FOR numb, Y ? Gset and Y != X
                        PCC[X][Y] = calculate PCC between Grating[X] and Grating[Y]
                        PCC_Ordered[X] = add social circle Y and corresponding PCC value
                ENDFOR
                // select the top similar 15 social cirles for every social circle
                FOR N, i < SEL
                        SimilarSocialCircles[X][i] = owner group of PCC_Ordered[i]
                ENDFOR
        ENDFOR
```

Figure 5.11: Recommendation Data Set Formation for The Similar Social Circles Algorithm - Part 1

In the Figure 5.11, the beginning part of the script is included. The script starts with the evaluation of the PCC values between every social circle pair. After this estimation, the 15 social circles that have the highest PCC values are selected for every user. Also the owners of the selected social circles are indicated.

In the Figure 5.12, the second part of the script is included. In this part, the recommendation data set is formed with the similar social circles and their corresponding owner users.

Through the implementation of the study, different platforms are used in order to achieve the described structure. In this section, a detailed description of the implementation languages and their usage in the system architecture are described.

```
        // evaluation dataset
        Evaluation-Data UserAndSocialCircles[N][SELx2]
        FOR numb, Z ϵ Uset
                // include the selected similar social circles
                FOR N, j < SEL
                        UserAndSocailCircles[Z][j] = SimilarSocialCircles[Z][j]
                ENDFOR
                // include the owner users of selected social circles
                FOR N, j < SEL
                        UserAndSocailCircles[Z][SEL+j] =owner user of SimilarSocialCircles[Z][j]
                ENDFOR
        ENDFOR

END
```

Figure 5.12: Recommendation Data Set Formation for The Similar Social Circles Algorithm - Part 2

### 5.3.1 Application Creation on Facebook

Facebook is a social network environment that is developed by Facebook Inc., with the aim of allowing its users to communicate with their social environments [52]. Facebook serves a special development environment to its developers. In this platform developers are allowed to implement in Facebook specific languages and in some well known web languages.

Regardless from the development language it is strictly stated to obey the privacy rules of Facebook [55]. The accounts of the developers who develops an application that does not obey the privacy agreement, are banned by Facebook without stating a reason or time interval.

Facebook serves three different API libraries to its developers which are implemented in PHP, JavaScript and ActionScript languages. In the 'Suggest Me a Movie' application, it is decided to use PHP API library considering the PHP familiarity.

PHP Client API serves all the required functionalities for the validation of the application. Facebook forces all its applications to authenticate to the Facebook platform as the first step of the development process. The authentication is performed with an authentication key that is supplied uniquely by the Facebook itself. The key is specific to the applications and cannot be shared between other applications. With the usage of the API calls, the creation of the required authentication session of the applications is performed.

In addition to that, some specific API calls are used for the transmission of the user related data to the 'Suggest Me a Movie' application from the Facebook environment. Although

Facebook does not allow to save logged in users' Facebook account information considering the privacy statements, it allows to access to a limited part of it by FBML tags, after including the required libraries.

The user interface of Suggest Me a Movie is implemented in FBML and HTML languages. FBML is used for Facebook related interface objects such as tabs, frames and user list boxes. Moreover, the user related information, such as names and friend lists are integrated in the application with the usage of FBML tags. On the other hand, for the rest of the interface related issues are implemented in standard HTML and JavaScript.

In the back ground, the communication between 'Suggest Me a Movie' application and its database is implemented in PHP. All the item and user information is saved and retrieved with the functions that are implemented in PHP 5.

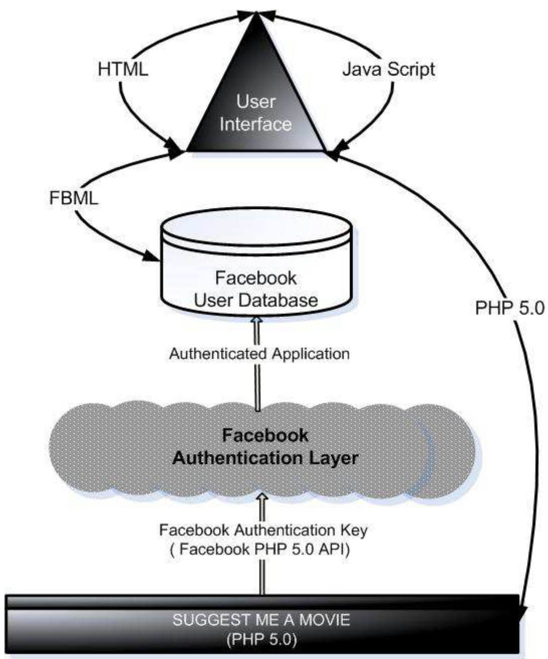


Figure 5.13: Implementation Architecture of Application

In the Figure 5.13, the general structure of the implementation environment of 'Suggest Me a Movie' is presented. As stated in the figure, firstly the application authenticates to the Facebook environment with the supplied authentication key. After this authentication process, Facebook shares its users' data with the application. In this study, the user interface is shaped with the usage of this personal data which is accessed by FBML tags. For the general interface issues, HTML and JavaScript languages are used.

### 5.3.2 Recommendation Agents

Recommendation Agents that perform the required calculations for generating the suggestions are developed independently from the Suggest Me a Movie application. This decision is made regarding to the fact that the integration of the agent execution to the application is impossible.

Considering large amount of the data set, since it is not applicable to generate recommendations after every user feedback, it is decided to execute the agents in a regularly period separately from the application. By this way the long execution time problem of the algorithms are solved without reflecting the side effects to the users. Also based on the assumption that one day is required to collect a meaningful data change, the execution interval of the algorithms is stated as one day.

For the implementation of the agents, Java is selected as the development language. The reasons of the usage of Java language can be listed as;

- Easy syntax that is driven from C and C++ languages.
- Simple object model and few low level facilities.
- Existence of Java Virtual Machine.
- Independency from the computer architecture and operating system.

In the Figure 5.14, the implementation architecture of the recommendation agents are indicated. In this study, two different recommendation agents are implemented which are Recommendation Generation Based on The Social Circles of Similar Users (stated as Similar Users) and Recommendation Generation Based on The Similar Social Circles (stated as Similar Social Circles). Both of the agents which are implemented in Java, recommend based on the data on Suggest Me a Movie database. The suggestions are shared with the users through the Facebook application.

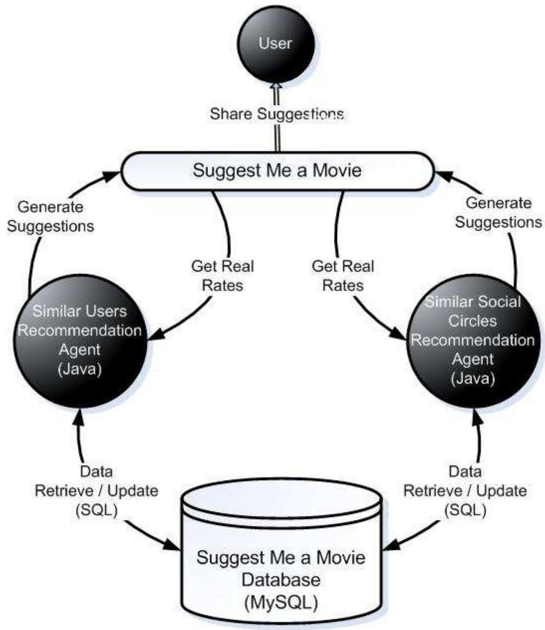


Figure 5.14: Implementation Architecture of Agents

### 5.3.3 Database

A database structure is used between 'Suggest Me a Movie' application and the recommendation agents with the aim of achieving communication between system components. A MySQL database is selected for the data storage as a free, fast and reliable open source relational database.

Although it is known that Oracle serves a more flexible and sophisticated platform for the developers, it is not preferred because of its complicated administrator issues.

# CHAPTER 6

# EVALUATION

## 6.1 Methodology

In the evaluation phase, it is aimed to identify the success of the studied algorithms. The suggestions of 'Recommendation Generation Based on The Social Circles of Similar Users' and 'Recommendation Generation Based on The Similar Social Circles' algorithms are shared with the users in order to retrieve their real opinions. In this part of the study, it is aimed to identify if the algorithms do generate concrete and accurate suggestions to the users. As an additional aim, it is tried to be identified which algorithm is better considering the accuracy of the recommendations. Beside users criticize if the recommendations are interesting enough. With the aim of producing more correct results, the evaluations have been done in two different strategies, statistical and user evaluation.

For the statistical evaluation the well known metrics are used for the calculation of the mean error between predicted ratings and actual ratings of the users. On the other hand, in the user evaluation process, the criticism of the recommendations is performed by the users themselves by answering the questions of an overall survey at the end of the suggestion evaluation process.

### 6.1.1 Data Retrieval

With the aim of retrieving data from the users, a new tab is created on the application that is named as 'My Suggestions'. In this section the generated recommendations are shared with the users for asking their real opinions. In this section, a very similar interface is used with the other parts of the application. For every suggested item, the comments that should be available are decided as;

- I watched it and liked it.
- I watched it but didn't like it.
- I didn't watch it. But I like to watch.
- I didn't watch it. But I don't like to watch.

On this section, it is decided to use the same comment list in order to not to confuse the users. Although ranking is not possible on the listed comments, since the data has been retrieved in the same way, it is thought that this will not decrease the accuracy of the estimations.

Moreover, in order to compare the studied two algorithms, the users are informed about the generation algorithm of the suggestions. With this aim, the suggestion methodologies of the recommended items are shared with the users. Every recommended item is served to the users with general information of its generation algorithm. Instead of a technical explanation of the algorithms, only symbolic names are indicated to not to distract their attentions.

Table 6.1: The Symbolic Names of Algorithms

| Symbolic Name | Algorithm |
|---|---|
| Similar Users | The Recommendation Generation Algorithm Based on The Data Retrieved from Social Circles of Similar Users. |
| Similar Social Circles | The Recommendation Generation Algorithm Based on The Data Retrieved from Similar Social Circles. |

In the Table 6.1, you can find the symbolic names of the used algorithms.

A screen shot is included that shows the general appearance of the 'My Suggestions' Section (Figure 6.1).

In the evaluation phase, it is decided to share 8 movies for every recommendation algorithm. The reason of this selection is not to bore the users. It is concluded that a small recommendation set should be shared with the users with the aim of supplying a completed data set considering the risk of non-complete data.

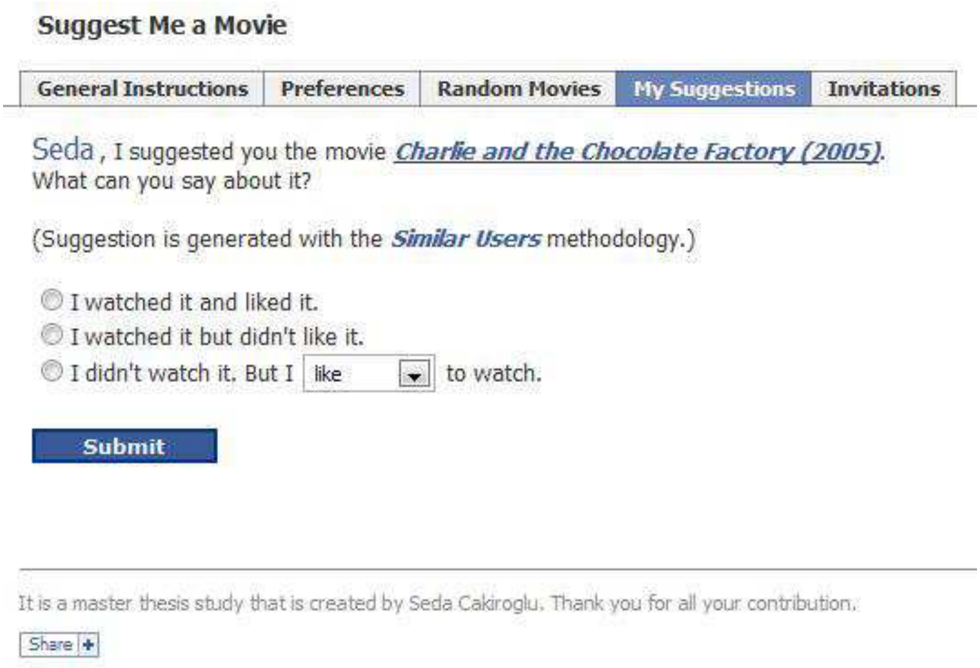


Figure 6.1: My Suggestion Section

The case of overlapping between the suggestion sets of these two algorithms are also considered. Although the accuracy of the evaluation of the algorithms are effected, the elimination of the repeated items is made on the suggestion sets. The repeated movie items are excluded from the algorithm's suggestion set in which the predicted rating is smaller.

Moreover, since a small overall survey is served to the users at the end of the suggestions, it is very critical to not to annoy them until the end of the personal evaluation part of the study. Regarding these facts, at total sixteen movie items are recommended to the users in this phase.

### 6.1.2 Metrics

#### 6.1.2.1 Statistical Evaluation Metrics

The statistical accuracy metrics, mean absolute error (MAE) and root mean squared error (RMSE) are used for the determination of the accuracy of the recommendation algorithms. The estimation results of both of the algorithms that are 'Similar User Methodology' and 'Similar Social Circle Methodology', are compared to each other in order to decide which one is better.

MAE [56] [25] is defined as the average absolute difference between the predicted ratings and the actual user ratings. The metric is linear; therefore all the differences in the sample set are weighted equally in the average.

$$MAE = \frac{\sum_{k,m} |x_{k,m} - \widehat{x}_{k,m}|}{L} \tag{6.1}$$

Where,

- k is an user who evaluates the suggested movies.
- m is one of the suggested items that the user k does not evaluate in the cold-start phase.
- $x_{k,m}$ is the predicted rating that the user k is likely to give to the item m.
- $\widehat{x}_{k,m}$ is the real rate of the user k that gives to the item m in the suggestion evaluation phase.
- L is the number of the suggested items that are evaluated by the users.

RMSE [26] [23] [56] is a quadratic equation in which the sample set ratings are squared and averaged. With the usage of this metric, the large differences in the predicted ratings and actual user ratings are weighted more on the average.

$$RMSE = \frac{\sqrt{\sum_{k,m} |x_{k,m} - \widehat{x}_{k,m}|^2}}{L} \tag{6.2}$$

Where,

- k is an user who evaluates the suggested movies.
- m is one of the suggested items that the user k does not evaluate in the cold-start phase.
- $x_{k,m}$ is the predicted rating that the user k is likely to give the item m.
- $\widehat{x}_{k,m}$ is the real rate of the user k that gives to the item m in the suggestion evaluation phase.
- L is the number of the suggested items that are evaluated by the users.

#### 6.1.2.2 User Evaluation Metrics

Near the statistical evaluation metrics, it is decided to include an additional survey at the end of the evaluation phase for retrieving the users' ideas. With this aim a very short overall questionnaire is served to the users. The questions that ask for recommendation algorithms' success are included in the survey.

Table 6.2: The Survey

| Algorithm Name | Feature | Options |
|---|---|---|
| Similar Users Methodology | Correct | yes |
| | | no |
| | Innovative | yes |
| | | no |
| Similar Social Circles Methodology | Correct | yes |
| | | no |
| | Innovative | yes |
| | | no |

In the Table 6.2, the questions that are included in the survey are showed. For both of the recommendation algorithms the features, the correctness and innovation are asked. The correctness criteria is considered as the accuracy of the suggestions. The innovation criteria is used for indicating if the recommendations are interesting and unexpected enough for the users.

Considering the possibility that the users cannot distinguish the recommendations based on their recommendation algorithms, the lists of the movies are included again at the beginning of the overall survey. For every algorithm, the items are remembered again by displaying their names in a list. In the Figure 6.2, the screen shot is included.

### 6.1.3 Expected and Actual Domain Parameters

In the cold start of the application when no data is available about the personal and social circle preferences of the users, it is decided to present the random movie items to the users. By the feedbacks that they supply it is aimed to recognize them for the recommendation process. In this phase, the expected number of users is predicted as 40. Since it is a small

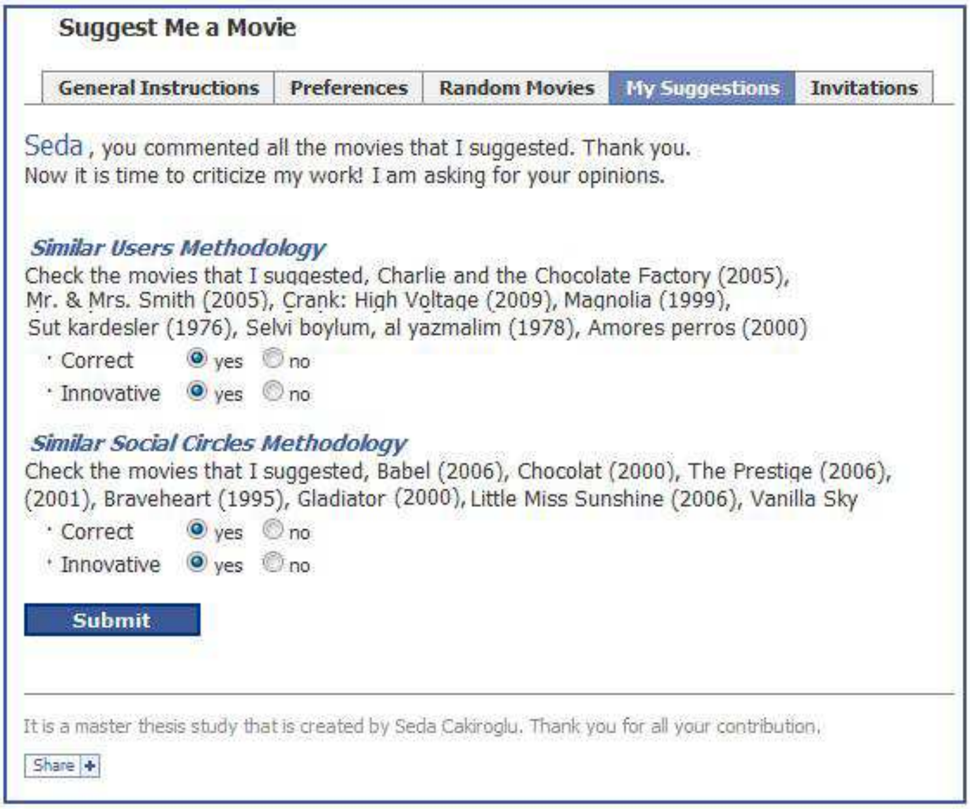


Figure 6.2: Overall Survey

spectrum study, 40 seems as a meaningful number for the test users that supply meaningful and concrete information.

For every user, it is asked to comment on the randomly selected movie items. Although there is no limitation, it is expected to comment on 40 movies at least in order to supply a meaningful and concrete data for the algorithms. Considering these parameters it is decided to use a movie data set that has 310 items for ensuring that the intersection is created between commented movie sets of these 40 users.

Table 6.3: The Expected and Actual Domain Parameters

| Domain Parameters | Expected | Actual |
|---|---|---|
| User | 40 | 52 |
| Movie Item | 310 | 310 |
| Commented Movie (Avg.) | 40 | 44 |

At the end of the cold start phase, it is identified that the number of the users that commented more than 40 movies is 52 (Table 6.3) which is better than the expected value. The average number of the movies that are commented by these 52 users is evaluated as 44. With this estimation, it is understood that %86 of the item set is evaluated based on the data retrieved. Additionally, with the consideration of that 16 movies will be suggested to the users at the evaluation process of the algorithms, it is concluded that only the top %5 of the evaluated items will be shared to the users as the recommendations.

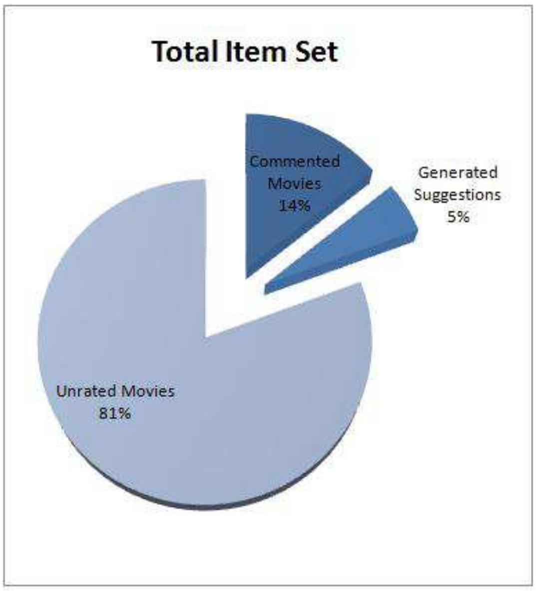


Figure 6.3: Total Item Set

In the Figure 6.3, the average percentages of the commented items, generated suggested items and unrated items are displayed.

#### 6.1.3.1 Statistical Usage Data of The Application

Facebook shares the usage statistics of the application with the developers. As a Facebook advantage, there is the possibility of accessing the tracked data of the usage and http request and allocations and user response and user profile information. Although through this study the user profile and active user counts are tried to be identified statistically, it is required to announce the Facebook statistics considering the accuracy and the application trust.

Table 6.4: The Users Statistics of Facebook

| Status | Number |
|---|---|
| Monthly Active Users | 5 |
| Total Users | 128 |
| Published Wall Post | 1 |
| Reviews | 4 |

In the Table 6.4, the total number of the users and activity items are shared with the aim of giving general information about the application popularity. As it is stated although the monthly active users are just 5, it is identified that whole number of the users who authenticated the application is 128. Wall Post is an action in Facebook which allows sending messages to the profile page of the applications. For 'Suggest Me a Movie' application it is indicated that only 1 wall-post is available. Besides this only 4 of these 128 users shared their opinions as review messages on the application profile page.

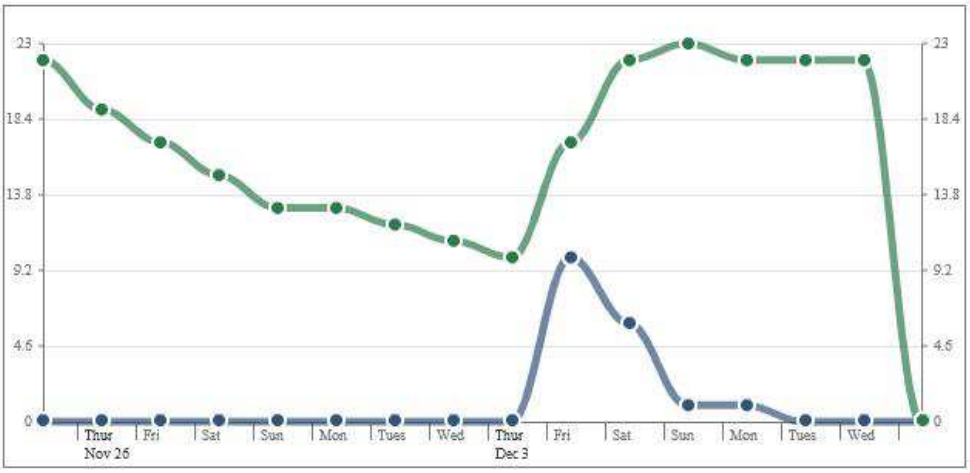


Figure 6.4: Application Usage Distribution

In addition to this, the usage statistics of the application is stated (Figure 6.4). In the graph the number of users who use the application are showed.

### 6.1.4 Evaluation Results

#### 6.1.4.1 Statistical Evaluation Results

With the aim of identifying the success of the algorithms, MAE and RMSE values are separately calculated for both of them. As a result of the feedback that supplied through the comments of the users to the recommended items, the calculated predicted ranks are compared to the actual ranks.

Table 6.5: Statistical Evaluation Results

| Algorithm | MAE | RMSE |
|---|---|---|
| Recommendation Generation Based on Social Circles of Similar Users | 3.76 | 0.39 |
| Recommendation Generation Based on Similar Social Circles | 2.90 | 0.28 |

As stated in the Table 6.5, for the 'Recommendation Generation Based on The Social Circles of the Similar Users' algorithm, MAE value is estimated at 3.76 and RMSE value is calculated as 0.39. On the other hand, for the algorithm 'Recommendation Generation Based on The Similar Social Circle', MAE value is estimated as 2.90 where RMSE value is calculated as 0.28.

It is concluded that similar social circle algorithm generates more successful suggestions compared to the social circles of the similar users algorithm. Also regarding to the small RMSE values, it is accurate to say that the predictions are very close to the real ranks that are supplied by the users.

Considering these results, it is accurate to say that the algorithms are very successful to generate the recommendations. It can be referred that the social circles of the users supply very accurate data for the recommendation algorithms and can be used in the further studies.

#### 6.1.4.2 User Evaluation Results

Based on the user comments that are supplied through the overall evaluation survey, it is concluded that the users are very satisfied by the recommendations. As stated below, most of the users supply very positive feedback for both of algorithms.

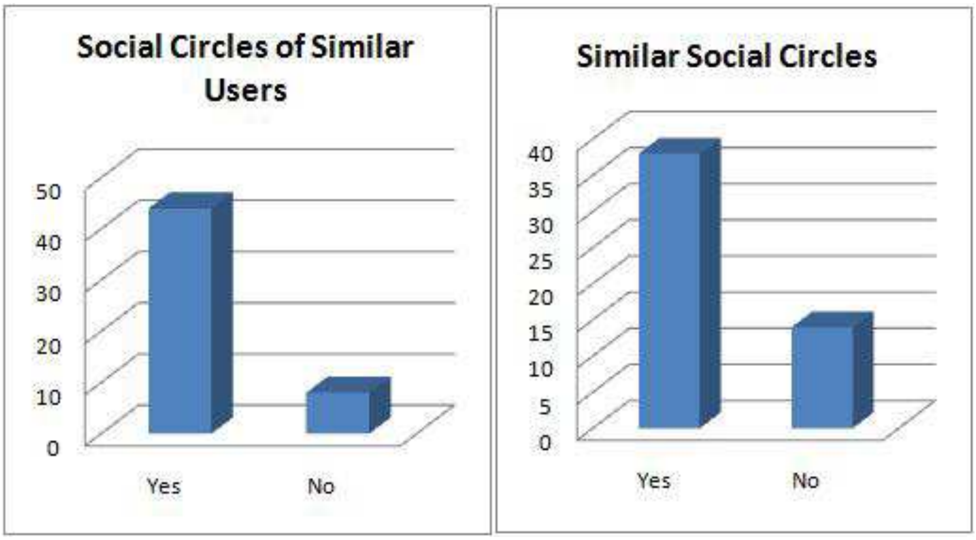


Figure 6.5: Feedback of The Recommendations

As stated in the Figure 6.5, it is identified that the recommendations that are generated through 'Social Circles of The Similar Users' algorithm generates more correct suggestions comparing to the 'Similar Social Circles'algorithm. 44 of 52 users say correct for the recommendations of social circle of similar user algorithm where 38 users says correct for the recommendations of the similar social circle algorithm.

On the other hand, as stated in the Figure 6.6 it is identified that the users find 'Similar Social Circles' algorithm more innovative comparing to the 'Social Circles of Similar Users' algorithm. All of the users state that the similar social circle algorithm's recommendations are interesting and unexpected where 46 of 52 users says that the recommendations of social circles of similar users algorithm's recommendations are interesting and unexpected.

#### 6.1.4.3 Overall View

With the accomplishment of the evaluation process, it is observed that the results are very different from the expected ones. Based on the statistical and personal evaluations it is concluded that suggestions of Similar Social Circles algorithm is better than the suggestions of the Social Circles of Similar Users algorithm.

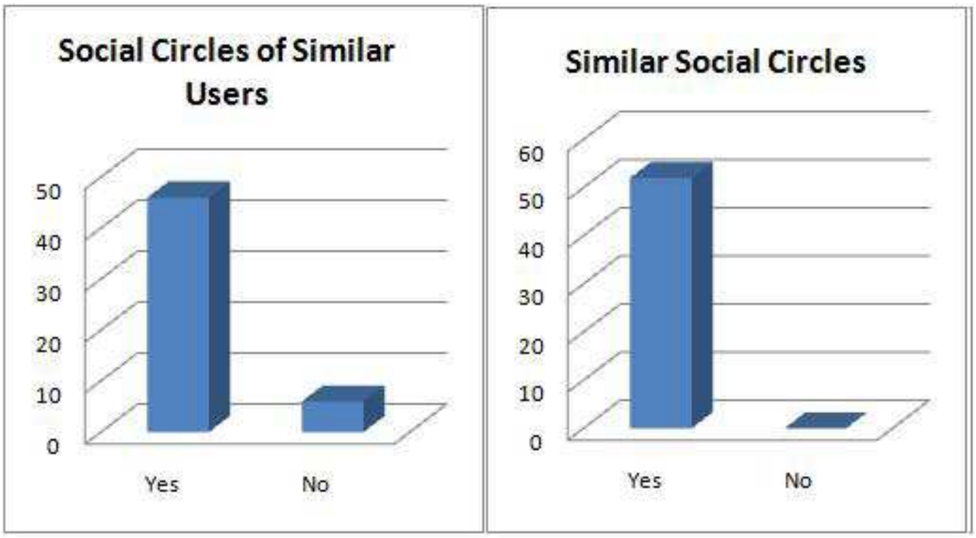


Figure 6.6: Innovation of The Recommendations

Table 6.6: Question #1

| | |
|---|---|
| Question#1 | if two people are alike to each other, does that also mean that their social environments are also alike at an acceptable level? |
| Description | If the movie preferences of two people are alike to each other, is it accurate to say that the common preferences of their social environments are also alike? Considering the fact that suggestions can be generated from social circles of the users, is it possible to use common preferences of these similar social circles for suggestions? |

As stated in the beginning of this study, it has been aimed to answer two questions in the recommendation area. For reminding the questions, they are listed in the Tables 6.6 and 6.7. Before the accomplishment of the study, the expectation was identifying the Social Circles of Similar Users algorithm as more correct compared to the Similar Social Circles algorithm. For the Similar Social Circles methodology, more interesting suggestions was expected, however the correctness had not been considered as the main concern.

The personal results show that although the Social Circles of Similar Methodology generates the suggestions with higher correctness rate, they are worse than Similar Social Circles algo-

Table 6.7: Question #2

| | |
|---|---|
| Question#2 | If social environments of two people are alike to each other, does that also mean that they are alike at an acceptable level? |
| Description | If common movie preferences of social groups of two different people are alike to each other regardless from their personal similarity, is it accurate to say that their movie preferences are also alike? Although these two people, that have similar social circles, are not alike to each other, is it possible to use preferences of one of them for suggestions to other one? |

rithm based on the unexpectedness of the suggestions. Besides, the statistical evaluations also figure out that in the overall perspective, the Similar Social Circles algorithm is better than the Social Circles of the Similar Users algorithm.

Table 6.8: Answer of Question #1

| | |
|---|---|
| Question#1 | If two people are alike to each other,does that also mean that their social environments are also alike at an acceptable level? |
| Answers | The results show that if two people are alike to each other, they also have alike common preferences in their social circles. The similarity rate of the social circles is at acceptable level for the recommendation algorithms. The correctness of the recommendations are approved by this study. |

In the tables 6.8 and 6.9, the answers of the stated questions are indicated. Considering all the phases of the study, it is concluded that the study achieves its objectives.

#### 6.1.4.4 Comparison with Other Recommendation Algorithms

Although, the methodology is not exactly the same, the study of Jung can be declared with the aim of comparing the study results [59]. In Jung's study, the visualization of recommender system on social network is studied by applying Friend of Friend methodology on MovieLens

Table 6.9: Answer of Question #2

| | |
|---|---|
| Question#2 | If social environments of two people are alike to each other, does that also mean that they are alike at an acceptable level? |
| Answer | The results show that if the social circles of two users are alike to each other, their social circles can be used to generated suggestions to each other. The results have showed that the similarity rate of users is at acceptable level for the recommendations. Additionally, the innovation of the algorithm are approved. |

data set.

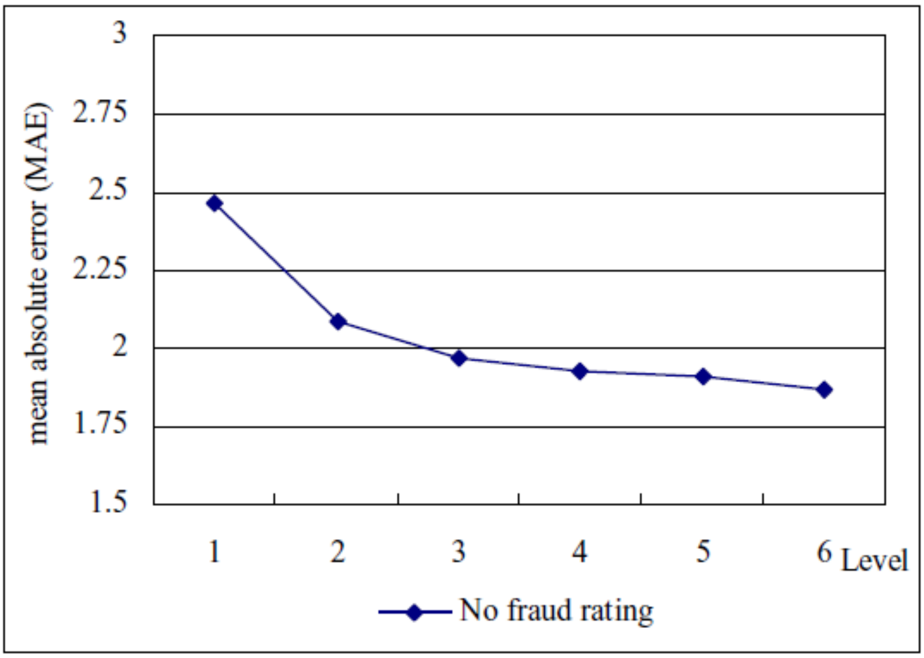


Figure 6.7: System Performance with Malicious Ratings

In the Figure 6.7, the comparison results of the Jung's study is stated. In the figure MAE values versus algorithm execution levels are indicated The experimental studies show that MAE value range is 2.5-1.75. In the figure, it is clearly shown that as the execution level increases the MAE is decreasing.

Although it was better from the current study results, it is necessary to remark that the huge data set of Jung's study, which includes 450 users with 66926 ratings, is an advantage for generating more accurate suggestions to the current study which includes only 52 users and 2289 ratings.

# CHAPTER 7

# DISCUSSION

Through this study, several problems have been experienced and with the aim of accomplishment of the study, various assumptions are made in order to solve some of the problems. Regarding to the study phases and unfamiliar development environment Facebook, there have been unknown difficulties besides the well known recommendation system's troubles. At the architectural design phases of the study, these difficulties were identified and best action items for each of them were selected. The rest of the study was tried to be built considering these action items.

## 7.1 The Difficulties Encountered and Their Solution Approaches

### 7.1.1 Irrelevant Friends on Facebook Account

Facebook is the largest and the most popular social network of the new generation. Considering this extensive huge community data, it is accurate to say that most of the users' friend lists include some unrelated people that they have not got any relationship and similarity in reality. By using the simple architecture of Facebook, the users are becoming friends of unrelated peoples and connect unfamiliar networks with the aim of enlarging their friend lists. As stated in the User Profile section, the number of friend of the application users is between 250-1000. By looking this huge numbers, we can infer that inspiring from all of these friends is not useful for a user in the movie domain.

For this study, since the social circle information is used as a representation of the user profiles, these unrelated friends create a very serious thread for the investigated aspects of the algorithms. In the case of using all of the friend lists of the users for the identification of the user's common movie tastes, the actual data about the users' preferences were messed up and the suggestions quality decreased.

After the identification of this problem, it is decided to use only a limited part of the friend lists of the users. Since there is no available criterion and data that helps me to make this elimination, it is concluded to allow users to perform this pruning by themselves. The personal referring opportunity is added to the application for allowing users to select their most important friends that affect their personal movie tastes.

Based on the assumption that if a user remembers a critic of his/her friend, that means this person shares some common movie taste with the user, the referring to the friend names is allowed in the social circle specific comment list. By this way, it is aimed to collect the close relationship information of the users.

With this strategy, the most informative friend lists of the users are structured for the rest of the study. By this way, it is ensured that the social circles of the users are generated in an accurate manner and they reflect the common movie preferences of the application users.

### 7.1.2 Intersection Between Users Friend-Lists

After the publication of the application, it was announced in my friend list in Facebook. Since the study users are commonly my friends, it is clear that most of the application users are friends with each other. That means for most of the application users, the friend lists are intersected with each other. The intersection can be seen as a thread for the success of the recommendation algorithms. Since considerable number of friends is common in the friend lists of the application users, there is the possibility of causing a wrong similarity decision between the social circles.

Considering the stated thread, in order to decrease the impact of the intersection between friend lists, the elimination of the social circles is applied as a solution. With the personal referring, the intersection area is tried to be minimized. Although it is still possible that two persons refer the same person in the application, since it shows that they have similar movie tastes, that should not damage the accuracy of the algorithms.

### 7.1.3 Long Execution Time of Suggestion Generation

As a common difficulty of the recommendation algorithms, since they are executed on huge data sets, their execution times are considerably high. Although several attempts are performed for improving the efficiency of the recommendation algorithms, such as applying dimensionality reduction technology [57], the execution time is still longer than expected.

The same situation is also applicable for the recommendation algorithms of the 'Suggest Me a Movie' application.

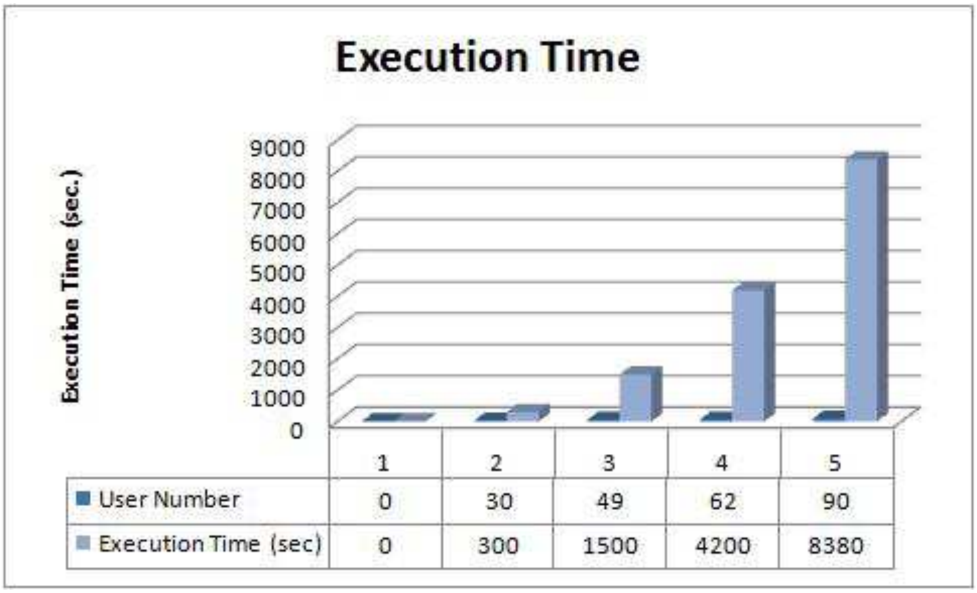


Figure 7.1: The Total Execution Time of The Algorithms

In the figure 7.1, the total execution time of the algorithms on different size of data sets are indicated. As indicated in the figure, the execution time increases exponentially based on the user number. Under these circumstances it is clear that the regeneration of recommendations after every submitted user comment is not possible. Regarding that it is decided to generate recommendations periodically.

For that reason, at every night, the algorithms are executed and the suggestions are updated according to the execution results. Besides considering that the data retrieved in a period shorter than one day, is not enough to create a meaningful change on the recommendations, one day is seemed as a reasonable interval to execute the algorithms.

### 7.1.4 Limited Data of Users in Cold-Start

Similarly with the collaborative filtering recommendation systems, 'Suggest me a Movie' also suffers from the 'first rater' and 'sparsity' problems [1]. Since the suggestions are generated for the users considering their existing comments on the movie items, no suggestions can be generated for the users that do not supply enough personal information about their movie tastes.

The stated difficulty is tried to be overcame by the usage of the referred friend information. By this way, although users do not supply enough data about themselves, by the personal and social data of the referred friends, the recommendation generation can be performed in cold-start phase of the application.

# CHAPTER 8

# CONCLUSION AND FUTURE WORK

In this study, two new recommendation algorithms that consider the users with their social circles are analyzed and evaluated. The study evaluation data is collected through a social network application, 'Suggest Me a Movie' on Facebook. Two different recommendation algorithms, which are 'Recommendation Generation Based on The Social Circles of Similar Users' and 'Recommendation Generation Based on The Similar Social Circles', are implemented and compared to each other.

In 'Recommendation Generation Based on The Social Circles of Similar Users' algorithm, the similar users are identified based on the user feedbacks. After this identification the social circles of the selected most similar users are used for the generation of the recommendations. In addition to this, the social circles of the indicated friends are also considered in the suggestion generation phase.

In 'Recommendation Generation Based on The Similar Social Circles' algorithm, the similar social circles of the users are identified based on the social circle feedbacks of the users. After this identification, the social circles and their owner users are used for the generation of the recommendations.

After gathering data from users in the cold start phase of the application, the suggestions are generated in a user specific way and shared with them through the 'Suggest Me a Movie' application. Based on the comments that the users supply for the suggested movies, the statistical evaluations are performed. In addition, an overall survey is served to the users in order to retrieve their opinions.

In the end of the evaluation phase, statistical results show that both of the algorithms are successful at the recommendation generation according to small MAE and RMSE values. In addition to these, it is identified that the social circles of the similar users algorithm generates more accurate suggestions whereas the similar social circles algorithm generates more interesting ones.

As a new aspect, a hybrid algorithm that includes 'Social Circles of The Similar Users' and 'Similar Social Circles' algorithms can be investigated. A sensible ratio between these algorithms can be evaluated with the aim of generating an optimum combination of the algorithms that has a higher accuracy and innovation rate.

In addition to these, in order to extend social circle idea, grouping of the friends based on the different parameters can be applied. By this way, the effects of different social circles can be investigated and the most effective social environments can be identified. The parameters that can be used for the social grouping can be listed as: age, gender and meeting way.

# REFERENCES


[1] Good, N., Schafer, B. J., Konstan, J. A., Borchers, A., Sarwar, B. M., Herlocker, J. L., Riedl, J., *Combining collaborative filtering with personal agents for better recommendations.* In: AAAI/IAAI. pp. 439-446, 1999.

[2] Belkin, N. J., Croft, W. B., *Information filtering and information retrieval: two sides of the same coin?* Commun. ACM 35, 29-38, 1992.

[3] Google, http://www.google.com, Last accessed on 13th March 2010.

[4] Kleinberg, J. M., *Authoritative sources in a hyperlinked environment.* Journal of the ACM 46 (5), 604-632, 1999.

[5] The Internet Movie Database, http://www.imdb.com, Last accessed on 2nd January 2010.

[6] Maes, P.,*Agents that reduce work and information overload.* Commun. ACM 37, 30-40, 1994.

[7] Lieberman, H., *Autonomous interface agents.* In: CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press, New York, NY, USA, pp. 67-74, 1997.

[8] Cohen, W. W., *Learning rules that classify e-mail.* In: In Papers from the AAAI Spring Symposium on Machine Learning in Information Access. pp. 18-25, 1996.

[9] Boone, G.,*Concept features in re:agent, an intelligent email agent.* In: AGENTS '98: Proceedings of the second international conference on Autonomous agents. ACM, New York, NY, USA, pp. 141-148, 1998.

[10] Resnick, P., Varian, H. R., *Recommender systems.* Commun. ACM 40 (3), 56-58, 1997.

[11] Oard, D. W., *The state of the art in text filtering.* User Modeling and User-Adapted Interaction 7, 141-178, 1997.

[12] Burke, R., *The wasabi personal shopper: A case-based recommender system.* In: The 11 Annual Conference on Innovative Applications of Artificial Intelligence, 1999.

[13] Schafer, B. J., Konstan, J. A., Riedi, J., *Recommender systems in e-commerce.* In: ACM Conference on Electronic Commerce. pp. 158-166, 1999.

[14] Nichols, D., *Implicit rating and filtering.* In: Proceedings of 5th DELOS Workshop on Filtering and Collaborative Filtering. ERCIM, pp. 31-36, 1998.

[15] Stevens, C., *Knowledge-based assistance for accessing large, poorly structured information spaces.* Ph.D. dissertation, Deptartment of Computer Science, Univ. of Colorado, Boulder, 1993.

[16] Adomavicius, G., Tuzhilin, A., *Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions.* Knowledge and Data Engineering, IEEE Transactions on 17, 734-749, 2005.

[17] Melville, P., Mooney, R. J., Nagarajan, R., *Content-boosted collaborative filtering for improved recommendations.* In: Eighteenth national conference on Artificial intelligence. American Association for Artificial Intelligence, Menlo Park, CA, USA, pp. 187-192, 2002.

[18] Grant, S., McCalla, G. I. ,2001. *A hybrid approach to making recommendations and its application to the movie domain.* In Proceedings of the Canadian AI Conference, pp. 257-266, 2001.

[19] Morita, M., Shinoda, Y., *Information filtering based on user behavior analysis and best match text retrieval.* In: SIGIR '94: Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval. Springer-Verlag New York, Inc., New York, NY, USA, pp. 272-281, 1994.

[20] Tang, T. Y., Winoto, P., Chan, K. C. C., *On the temporal analysis for improved hybrid recommendations.* In: WI '03: Proceedings of the 2003 IEEE/WIC International Conference on Web Intelligence. IEEE Computer Society, Washington, DC, USA, pp. 214+, 2003.

[21] Berkovsky, S., Kuflik, T., Ricci, F., *Cross-representation mediation of user models.* User Modeling and User-Adapted Interaction 19 (1-2), 35-63, 2009.

[22] Goldberg, D., Nichols, D., Oki, B. M., Terry, D., *Using collaborative filtering to weave an information tapestry.* Communications of the ACM 35, 61-70, 1992.

[23] Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R., Riedl, J., *Grouplens: Applying collaborative filtering to Usenet news.* Communications of the ACM 40 (3), 77-87, 1997.

[24] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J., *Grouplens: an open architecture for collaborative filtering of netnews.* In: CSCW '94: Proceedings of the 1994 ACM conference on Computer supported cooperative work. ACM, New York, NY, USA, pp. 175-186, 1994.

[25] Shardanand, U., Maes, P., *Social information filtering: Algorithms for automating "word of mouth".* In: Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems. Vol. 1. pp. 210-217, 1995.

[26] Hill, W., Stead, L., Rosenstein, M., Furnas, G., *Recommending and evaluating choices in a virtual community of use.* In: CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM Press Addison-Wesley Publishing Co., New York, NY, USA, pp. 194-201, 1995.

[27] O'Mahony M., Hurley N., Kushmerick N., and Silvestre G., *Collaborative recommendation: A robustness analysis.* In Proceedings of Int'l Semantic Web Conf., 2003.

[28] Golbeck J., Hendler J., and Parsia B., *Trust networks on the Semantic Web.* In Proceedings of Cooperative Intelligent Agents, 2003.

[29] Massa, P., Avesani, P., *Trust-aware collaborative filtering for recommender systems.*, 2004.

[30] P. Bonhard, et al., *Accounting for taste: using profile similarity to improve recommender systems.* In CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems, pp. 1057-1066, New York, NY, USA. ACM, 2006.

[31] Liu, H., *Social network profiles as taste performances.* Journal of Computer-Mediated Communication 13 (1), 252-275, 2007.

[32] Terveen, L., Hill, W., *Beyond Recommender Systems: Helping People Help Each Other.* Addison Wesley, pp. 487-509, 2001.

[33] A.M. Collins, and E.F. Loftus, *A spreading-activation theory of semantic processing.* Psychological Review, 82, pp. 407-428, 1975.

[34] Donath, J., Boyd, D., *Public displays of connection*. BT Technology Journal 22 (4), 71-82, 2004.

[35] Ackerman, M. S., *Augmenting organizational memory: a field study of answer garden.* ACM Trans. Inf. Syst. 16, 203-224, 1998.

[36] Maltz, D., Ehrlich, K., *Pointing the way: active collaborative filtering.* In: CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, pp. 202-209, 1995.

[37] Sinha, R., Swearingen, K., *Comparing recommendations made by online systems and friends.* In: In Proceedings of the DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries, 2001.

[38] Sifry D., *State of the blogosphere:Part 1 - on blogosphere growth.* http://technorati.com/weblog/2006/04/96.html, 2006.

[39] Orkut, http://www.orkut.com, Last accessed on 4th November 2009.

[40] MySpace, http://www.myspace.com, Last accessed on 3rd January 2010.

[41] Facebook, http://www.facebook.com, Last accessed on 3rd March 2010.

[42] Wheeless L. and Grotz J., *The Measurement of Trust and Its Relationship to Self-disclosure.* Communication Research 3, pp. 250-257, 1977.

[43] Herlocker J., Konstan J. and Riedl J., *Explaining Collaborative Filtering Recommendations.* Conference on Computer Supported Cooperative Work, pp. 241-250, 2000.

[44] Wikipedia, http://en.wikipedia.org/wiki/Last.fm, Last accessed on 3rd March 2010.

[45] Eck, D., Lamere, P., Bertin-Mahieux, T., Green, S., *Automatic generation of social tags for music recommendation.* 2007.

[46] Firan, C. S., Nejdl, W., Paiu, R., *The benefit of using tag-based profiles.* In: LA-WEB '07: Proceedings of the 2007 Latin American Web Conference. IEEE Computer Society, Washington, DC, USA, pp. 32-4, 2007.

[47] Friend of a Friend, http://www.foaf-project.org/, Last accessed on 5th March 2010.

[48] IBM Developers, http://www.ibm.com/developerworks/xml/library/x-foaf.html, Last accessed on 8th March 2010.

[49] Suggest Me a Movie on Facebook, http://apps.facebook.com/suggestmeamovie, Last accessed on 8th March 2010.

[50] Oard, D., Kim, J., *Implicit feedback for recommender systems.* In: in Proceedings of the AAAI Workshop on Recommender Systems. pp. 81-83, 1998.

[51] Sarwar, B. M., Karypis, G., Konstan, J. A., Reidl, J., *Item-based collaborative filtering recommendation algorithms.* In: World Wide Web. pp. 285-295, 2001.

[52] Facebook on Wikipedia, http://en.wikipedia.org/wiki/Facebook, Last accessed on 23rd January 2010.

[53] PCC on Wikipedia, http://en.wikipedia.org/wiki/Pearsoncorrelationcoefficient, Last accessed on 2nd February 2010.

[54] Papagelis, M., and Plexousakis, D., *Qualitative analysis of user-based and item-based prediction algorithms for recommendation agents.* In: CIA 04, 9th International Workshop on Cooperative Information Agents, 2004.

[55] Facebook Privacy Aggrement, http://www.facebook.com/policy.php, Last accessed on 11th February 2010.

[56] Sarwar, B. M., Konstan, J. A., Borchers, A., Herlocker, J. L., Miller, B. N., Riedl, J., *Using filtering agents to improve prediction quality in the grouplens research collaborative filtering system.* In: Computer Supported Cooperative Work. pp. 345-354, 1998.

[57] Berry, M. W., Dumais, S. T., O'Brien, G. W., *Using linear algebra for intelligent information retrieval.* Tech. Rep. UT-CS-94-270, 1994.

[58] Terveen, L., Hill, W., Amento, B., Mcdonald, D., Creter, J., *Phoaks: A system for sharing recommendations.* Communications of the ACM 40 (3), 59-62, 1997.

[59] Jung J., *Visualizing Recommendation Flow on Social Network.* J-Jugs, 1780-1791, 2005.

# APPENDIX A

# PUBLICATIONS

Seda ÇAKIROĞLU, Ayşenur BİRTÜRK, Suggest Me a Movie: A Multi-Client Movie Recommendation Application on Facebook, In 6th International Conference on Web Information Systems and Technologies (WEBIST'10) - Accepted as Short Paper but not registered.

Seda ÇAKIROĞLU, Ayşenur BİRTÜRK, Suggest Me a Movie: A Multi-Client Movie Recommendation Application on Facebook, In 25th International Symposium on Computer and Information Sciences (ISCIS'10) - Accepted as Full Paper.

Seda ÇAKIROĞLU, Ayşenur BİRTÜRK, Suggest Me a Movie: A Multi-Client Movie Recommendation Application on Facebook, to be submitted to the Computer Journal by July 2010 - In Preparation.