Islamic University of Technology
Department of Computer Science and Engineering (CSE)

Friend Recommendation System in Social Network using Personality Analysis and User Behavior

Authors
Nafis Neehal - 124407
&
Shoaib Bin Noor - 124423

Supervisor
Prof. Dr. M.A. Mottalib
Head, Department of CSE

Co-Supervisor
Md. Mohiuddin Khan
Assistant Professor, Department of CSE

A thesis submitted to the Department of CSE
in partial fulfillment of the requiremnts for the degree of B.Sc. Engineering in CSE
Academic Year: 2015-16

November - 2016

Declaration of Authorship

This is to certify that the work presented in this thesis is the outcome of the analysis and experiments carried out by Nafis Neehal and Shoaib Bin Noor under the supervision of Dr. M.A. Mottalib, Head of the Department of Computer Science and Engineering (CSE), Islamic University of Technology (IUT), Dhaka, Bangladesh. It is also declared that neither of this thesis nor any part of this thesis has been submitted anywhere else for any degree or diploma. Information derived from the published and unpublished work of others has been acknowledged in the text and a list of references is given.

Authors:

Nafis Neehal
Student ID - 124407

Shoaib Bin Noor
Student ID - 124423

Supervisor:

Dr. M.A. Mottalib
Head of the Department
Department of Computer Science and Engineering
Islamic University of Technology (IUT)

Acknowledgement

We would like to express our grateful appreciation for **Prof. Dr. M.A. Mottalib**, Head of the Department, Department of Computer Science & Technology, IUT for being our Supervisor.

Also we would like to thank Md. Mohiuddin Khan Rafi, Assistant Professor, Islamic University of Technology for all his efforts, guidelines and suggestions.

Their motivation, suggestions and insights for this thesis have been invaluable. Without their support and proper guidance this research would not have been possible. Their valuable opinion, time and input provided throughout the thesis work, from first phase of thesis topics introduction, subject selection, proposing algorithm, modification till the project implementation and finalization which helped us to do our thesis work in proper way. We are really grateful to then.

We would also like to thank Dr. Kamrul Hasan, Associate Professor, Department of CSE, IUT and Hasan Mahmud, Assistant Professor, Department of CSE, IUT for their invaluable suggestions and guidelines.

Abstract

Social networking is a tool used by people all around the world. Its purpose is to promote and aid communication. Social networks, such as Facebook, were created for the sole purpose of helping individuals communicate. These networks are becoming the modern way to make friends. These new friends communicate through these networks. There exist recommendation systems in all the social networks which help users to find new friends and connect to more peoples. With friends, there comes a strong friend recommendation system also. The existing social networks do have their own friend recommendation system which is based on the friends of friends' methodology. This graph based friend recommendation system is not very accurate most of the time and drive users to wrong direction. We tried to make this recommendation system more accurate adding some extra layers of personality analysis and user behavior. With the vast amount of user data, our system will figure out each user's personality traits and behavior which will be used to help him/her finding out new users with same nature.

Contents

1 Introduction

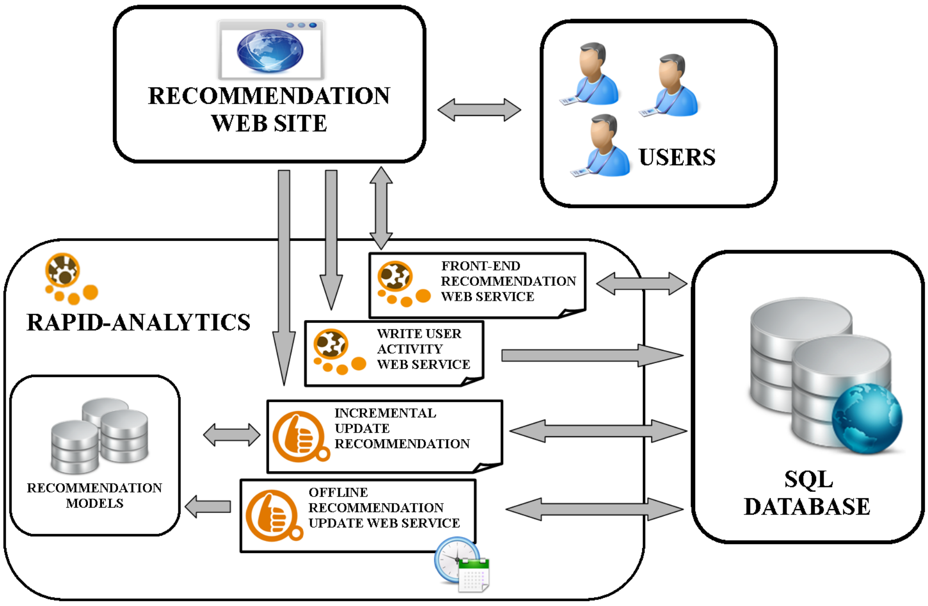
1.1 Overview

The number of people using the social network is increasing. Social networks today made our life easier by helping us get connected quickly with people all around the world saving time and money. Moreover, getting connected to more people allows us to know and get used to with different cultures. So, Social Networks without friends is just like an empty useless box. Entering an era of data explosion, the network has become an information sharing platform; it is increasingly difficult to find the information people need in such vast amounts of data. In the condition that people know what information they really need, the search engine provides users with a quick and convenient way to find appropriate items.

However, when facing the large amount of data in the network, there are many users who aren't aware of what they are really interested in or what they need. As a result, the recommendation engine provides a good way to deal with the problem and satisfies the users. Recommendation engine helps users filter out the apt information they may need indeed from the flood of information, recommending the users with items, according to the users' behaviors in the past and their information the systems collected.

Nowadays, the recommendation engine has been successfully applied to the field of electronic commerce and some music and movies websites, such as Amazon, Netflix, etc. With the rise of a variety of social networks such as micro blog and forums, any of them have launched an application which can recommend potential friends users may be interested in such as Facebook and twitter. While they both recommend information to the target users, the recommendation engine in the social network is of considerable different from the traditional recommendation system applied in the Electronic Commerce. Compared with the friend recommendation, the traditional-recommendation application has the characters that it has less attributes rarely changed in the items and that the impact factors on the users and items are relatively simple, for example, there can be properties like genre, and writer, singer and publication date in the music.

However, when considering the relations between people and people, there are many factors related. First, describing a person's interests and hobbies is more complex than the description of items. Secondly, people's interests and hobbies change over time. Thirdly, in the social network, for most of the users, they add people into their friend lists not based on similar interests, but due to geographical location, study or work experience. In this way, recommending potential friends that users may interested in would be influenced by factors above. Importantly, similar people users met through social network are much valuable for themselves. People often gain more help from these weak ties than from strong ties. So research on the social network friend-making method and potential friend recommendation is very necessary.



The idea used in PYMK Friend Recommendation designed by Myspace is based on the user's social relations and personal information profiles, or, the 'friend of friend' method. The similar strategy is applied on Facebook, Renren, twitter and other social networks, in which the recommended results are based on friends in common, similar age, geographic location, school information and etc. However, there are some drawbacks to extend friends through this way. Friends recommended based only on the number of common friends may not necessarily have the same hobbies and habits with the target users. Thus, the user may associate the so-called potential friends and may fall victim into the dilution of 'friendship' based on the above strategy, for the reason that they may share some contents that the target users may not interested. At the same time, this method may cause the waste of lots of useful information in social networks, through which we can obtain more accurate and meaningful friend's recommendation results.

1.2 Problem Statement

The primary goal of a friend recommendation system in social network is to provide with the most relevant data to the user based on their requirement or demand. But now-a-days in social networks there are too much data leading to an overwhelming condition. For instance, if we take Facebook, it has a worldwide monthly active user of 1.26 Billion which is increasing by 15% every year [2, 3, 5].

Also there is the factor that Facebook or any other recommendation system does not allow users to choose the category of people they want to be friends with. In Facebook, friends are recommended based on people a user searches for, people who searched for the user, number of mutual friends, group affiliation and so on.

And there is now way that users can personalize this recommendation criteria. So it becomes very hard to choose friends among this huge amount of people. Recommendation system helps people by narrowing down the choice domain. Different people have different agenda. That's why developing a general recommendation system by adopting traditional methods to satisfy everyone is difficult.

As there are already tons of standards for friend recommendation in social network, we chose to proceed with analysing user personality and based on that we wanted to recommend friends.

1.3 Motivation & Scopes

1.3.1 Importance of Recommendation System

A recommendation system helps the people to find out what he/she desires from a vast collection of entities belonging to the same categories. A well designed recommendation system can also improve the experience of the users of websites or other services.

1.3.2 Techniques of Recommendation System

Many current literatures evaluate different types of recommendation system technologies, and their real life applications introduces and classifies recommendation system into three main techniques -

- Content based recommendation system techniques
- Knowledge based recommendation systems
- Community based recommendation system

The solutions to the problem identified in the traditional recommendation systems could be developed by applying social network data in recommendation systems. Integration of social networks can theoretically improve the performance of current recommender systems.

- First, in terms of the prediction accuracy.
- Second, with friends' information in social networks, it is no longer necessary to find similar users by measuring their rating similarities.

1.3.3 Applications of Recommendation System

Examples of new recommendation applications include Amazon.com's item recommendations, movie recommendations (MovieLens), webpage recommendations (Google). Facebook provides features to recommend a list of people you may know.

1.3.4 Importance of Recommendation System in Social Network

Although many researchers have discussed the usefulness of social network based predictions, recommendation system in social networking area is still in its early phase describes non-social network based recommendation system such as Collaborative Filtering as traditional methodology, and discussed its flaws and weaknesses. The solutions to the problem identified in the traditional recommendation systems could be developed by applying social network data in recommendation systems. Integration of social networks can theoretically improve the performance of current recommender systems.

First, in terms of the prediction accuracy. Second, with friends' information in social networks, it is no longer necessary to find similar users by measuring their rating similarities. When people are friends, there are certain things in common among them. Therefore, the social network based recommendation system makes the community based recommendation technique more powerful and useful. In recent times, there have been opportunities for novel recommender applications on the social web that directly involve humans in a recommendation process, in which users make recommendations to other users. This is called crowd recommendations.

1.4 Research Challenges

Crawling through social network data is a huge work as there are tons of numeric and non-numeric values. Also determining the features based on which the personality analysis would be done is also tough. After analyzing personality, categorizing users into different personality categories and based on their personality the recommend them as friends is the most hard job. Also there are two methods for recommendation. In one method, the user gets to choose the category of people he/she wants to be friends with. And the other method is, people are autonomously recommended friends based on the similarity of their personality which is calculated based on the cosine similarity.

1.5 Thesis Outline

In Chapter 1 we have discussed our study in a precise and concise manner. Chapter 2 deals with the necessary literature review for our study and there development so far. In Chapter 3 we have stated the skeleton of our proposed method, proposed algorithm for personality analysis and also the flowchart to provide a detail insight of the working procedure of our proposed method **A**dvanced **A**gglomerative **C**lustering **T**echnique(AACT) using Manhattan Distance. Chapter 4 shows the results and comparative analysis of successful implementation of our proposed method. The final segment of this study contains all the references and credits used.

2 Recommendation System Analysis

2.1 Recommendation System

When confronted to overwhelming plethora of choices [2] and lacking specific domain knowledge, users need some assistance exploring or winnowing down the possibilities. Then, Recommendation system helps users in drawing conclusion about relevant particular based on the information provided by the items that user liked in the past.

There are two main methods in the recommended areas to generate recommended systems-

2.1.1 Contend based recommendation system

Content-based filtering method is based on the user's previous behavior, such as users' items ratings, browsing records of goods and their purchasing history, establishing the user model to describe their interests and behaviors. At the same time, the system establishes a model for each item to describe its characteristics. And then match the user models and item models by estimating how much they are related each other, recommending the items users may need. For example, in a music-sharing site, we can establish a user model to describe the users' interests and preferences by collecting user's previous listening history, search history and other information. And then to recommend music's which match users' preferences described by the user model.

2.1.2 Collaborative filtering based recommendation system

Collaborative filtering and content filtering is dissimilar. Collaborative filtering recommendation recommends items for people based on users who are alike with them. While in content-based filtering algorithm, we calculate the similarity between the items, in the collaborative filtering algorithm, we analyze relationships between users and interdependencies among products to identify new user-item associations. Using collaborative recommendation, one identifies users whose tastes are similar to those of the given user and recommends items they have liked. Both information filtering method have a wide application in the field of ecommerce.

A major advantage of collaborative filtering is that it can address data aspects that are often elusive and difficult to profile using content filtering and generally speaking, it is more accurate that content-based method. Collaborative filtering suffers from what is called the 'cold start' problem, due to its inability to address the system's new products and users. There are considerable differences between people recommendation and items recommendation and the performance of the two methods remains to be seen when applied into the friend recommendation field. The approach used in nowadays social network, 'friends of friends' method can be seen as a way of adopting the ideas in collaborative filtering.

As the preceding discussion, in the social networking, many of the user's friendship are not based on users' interests and hobbies, but by other factors, for example, one may become friend with others for the reason that they come from the same city or country. 'Friends of friends' recommending method may compel users to add some people with whom they are not similar and who may share information the user is not interested in. And new 'friends' may derive from these people, lowering the accuracy of recommendation.

2.2 Friend Recommendation System

In real life, people often resort to their friends in social networks to share personal opinions, interests and typically rely on recommendations of them before purchasing a product or consuming a service. Hence, social networking sites, as a solution, provide novel ways for users to meet, communicate, and share interests and also serve as a platform for incentives in business marketing via modeling consumer behavior (Chen and Qi 2011; Bonchi et al. 2011). As a matter of fact, SNSs are in urgent need of new applications which can both grow and retain users and in turn offer them with various products and services that they liked in past. FRS, one such application, explores user's friends network through the given user attitude estimation toward another from the evidence provided by his relationship with other members of surrounding network.

Several methodologies and mathematical models (Kleinberg 2001) have developed to show how people interact with one another and establish link on SNSs. In the following subsections, the contributions of various link prediction approaches are described encompassing two major categories: graph-based and CF-based approaches.

2.2.1 Graph-based approaches

The task of recommending people to a specific user [14] is same as of predicting new links in social networks (Nowell and Kleinberg 2004; Chen et al. 2009; Guy et al. 2009; Leskovec et al. 2010; Brzozowski and Romero 2011). Most existing approaches have used articulated social network structure while suggesting friends (Nowell and Kleinberg 2004; Quercia and Capra 2009). For instance, Nowell and Kleinberg (2004) made recommendation of friends by considering only the local features of graph and compared several local similarity measures, such as Common neighbor, Jaccard's coefficient, Adamic/Adar, Shortest path, and Katz coefficient in a large study on social network sites.

On the other side, there are global approaches that utilize the complete complex structure of network but would be hard to comply with, because of the huge size of networks. Symeonidis et al. (2010) have incorporated transitive node similarity into global graph features that captures adequately the missing local graph characteristics and enhance its performance. Some researchers have found that the problem of link prediction can also be solved just by considering the common interest that people share with others instead of gathering the complete social network information, which could be either impossible or very expensive (Patil 2009; Xie 2010; Scellato et al. 2011).

2.2.2 CF Based Approaches

Another body of research, combining CF and social networks (Liu and Lee 2009; Bonhard et al. 2007; Huang et al. 2005; Kautz et al. 1997) in finding friends or experts in specific fields, considered direct relationship between people[14]. For example, Referral web, an interactive tool, expands users' awareness by providing them a series of people or document in response to their queries, via social network information (Kautz et al. 1997). Some studies compared content similarity and social relationship based algorithms for recommending people in an enterprise social network and found that content similarity algorithms perform better at discovering new friends (Chen et al. 2009). Guy et al. (2010) aggregated social network information from various sources and highlighted the three classes of user similarity sources, people, things, and place that they share, which yield analogous impact on recommendations.

Later, an idea of strong correlation of users' similarities with social information, i.e. trust came into picture that might significantly help in expanding user's friends' network (Ziegler and Golbeck 2005). Liang and Li (2011) built a hybrid system to recommend folks incorporating user interest information (or user preferences) and social information together into CF and showed that it outperforms traditional CF-based systems (used only item preference) and similarity-based methods (i.e. Adamic Adar, Common neighbor, and Jacard coefficient). Despite many studies based on content similarity, there is a strong need to enlist the set of attributes that strongly reflect user's behavior. A weighted approach (Garcia and Amatriain 2010; Yin et al. 2010; Agarwal and Bharadwaj 2011) is designed around the ability to measure the impact of various features that would encourage a user to connect with others.

Some other methods have also used evolutionary algorithms to adapt the realism of weighted schemes which make them more robust (Ujjin and Bentley 2004; Silva et al. 2010; Agarwal and Bharadwaj 2011; Bobadilla et al. 2011; Naruchitparames et al. 2011). Recently, (Yang et al. 2011a) proposed a FIP model that aims to link up users with interested services and with other users who share common interests via information contained in interest networks and friendship networks. On the other hand, this model also helped to bridge the gap between CF and random walk and dragged them into a unified framework that serves both the purposes, i.e. service recommendations and friends' suggestions simultaneously. So far, people have focused only on the use of proximity measures (such as common friends, work, education, etc.) in CF for friend recommendations, but Bian and Holtzman (2011) have studied it in a different way through the use of personality matching with CF and have established that this approach ensures a higher amount of sustainability in friendship.

2.3 Summary of Renowned Algorithms for Friend Recommendation

Some renowned algorithms are [5] -

2.3.1 Content-based approach

1. Heuristic-Based RS
   - TF-IDF (information retrieval)
   - Clustering
2. Model-based RS
   - Bayesian Classifiers
   - Clustering
   - Decision trees
   - Artificial Neural Network

2.3.2 Collaborative-based approach

1. Heuristic-Based RS
   - Nearest neighbor (cosine, correlation)
   - Graph theory
   - Clustering
2. Model-based RS
   - Bayesian Network
   - Clustering
   - Probabilistic models
   - Linear regression
   - Artificial Neural Network

2.3.3 Hybrid approach

1. Heuristic-Based RS
   - Linear combination of predicted ratings
   - Various voting schemes
   - Incorporating one component as a part of the heuristic for the other
2. Model-based RS
   - Incorporating one component as a part of the model
   - Building one unifying model

2.4 Common Limitations of Recommendation Systems

The common limitations of recommendation system are [5] -

2.4.1 New User problem

A recommendation system has no information to make recommendations about a new user. This is also called cold start problem.

2.4.2 Sparsity Problem

Due to the large amount of items and users, it is natural that users will only have ratings on a few items that are most relevant to themselves. This leaves a large amount other items not rated or not having social contacts by the users.

2.4.3 Over-Specialization

This is a problem when the system can only recommend the items that the user already saw or those with high scores and the user is limited to being recommended to the items that are similar to those already rated.

2.4.4 Limited Content Analysis

This is similar to the New User Problem and many times we don't have enough information regarding the items.

3 Personality Analysis

3.1 Human Personality Insight

Personality can be defined in different ways, depending on whether we focus on the individual or on people in general.

If we focus on people in general, then we can define personality in terms of individual differences — that is, the range of different styles of thinking, feeling and acting.

Just as human beings can differ a great deal in terms of their physical traits (height, weight, hair, and so on), they also differ in terms of mental and behavioral traits. For example, some people are noticeably talkative and outgoing while others are noticeably quiet and reserved. Such differences and variations are seen everywhere throughout the human population.

If we focus on the personality of a specific individual, we can define it as that person's particular set of enduring dispositions or long-term tendencies to think, feel and act in particular ways.

Personality is usually broken into components called the Big Five, which are openness to experience, conscientiousness, extroversion, agreeableness, and neuroticism (or emotionality). These components are generally stable over time, and about half of the variance appears to be attributable to a person's genetics rather than the effects of one's environment.

3.2 International Personality Item Pool References

- The International Personality Item Pool is a public domain collection of items for use in personality tests.
- It is managed by the Oregon Research Institute.
- The pool contains 3,329 items
- A Scientific Collaborator for the Development of Advanced Measures of Personality and Other Individual Differences
- The growing popularity of the IPIP can be attributed to five factors -
  1. It is cost free
  2. Its items can be obtained instantaneously via the Internet
  3. It includes over 2000 items, all easily available for inspection
  4. Scoring keys for IPIP scales are provided
  5. Its items can be presented in any order, interspersed with other items, reworded, translated into other languages, and administered on the World Wide Web without asking permission of anyone

3.3 Computer VS Human Accuracy in case of Predicting Human Personality

Personality is defined as the coherent patterning of affect [10], behavior, cognition and desire over time and space, which are used to characterize unique individuals. There are several theories for personality traits in the literature but the most widely used personality traits model is the Big-5, five broad personality dimensions. It describes the human personality as a vector of five values corresponding to bipolar traits. This is a popular model among the language and computer science researchers and it has been used as a framework for both personality traits identification and simulations.

Judging others' personalities is an essential skill in successful social living, as personality is a key driver behind people's interactions, behaviors, and emotions.

This paper showed clearly that

- Computer predictions based on a generic digital footprint (Facebook Likes) are more accurate (r = 0.56) than those made by the participants' Facebook friends using a personality questionnaire (r = 0.49)
- Computer models show higher inter-judge agreement
- Computer personality judgments have higher external validity when predicting life outcomes such as substance use, political attitudes, and physical health

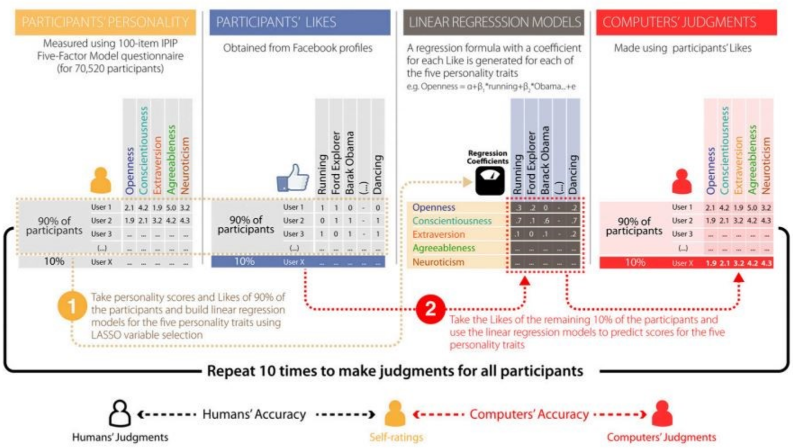


Figure 1: Class Diagram

According to the Realistic Accuracy Model, the accuracy of the personality judgment depends on the availability and the amount of the relevant behavioral information, along with the judges' ability to detect and use it correctly.

Such conceptualization reveals a couple of major advantages that computers have over humans -

- First, computers have the capacity to store a tremendous amount of information, which is difficult for humans to retain and access.
- Second, the way computers use information through statistical modeling generates consistent algorithms that optimize the judgmental accuracy, whereas humans are affected by various motivational biases.
- Nevertheless, human perceptions have the advantage of being flexible and able to capture many subconscious cues unavailable to machines. Because the Big Five personality traits only represent some aspects of human personality, human judgments might still be better at describing other traits that require subtle cognition or that are less evident in digital behavior.

This study is limited in that human judges could only describe the participants using a 10-item-long questionnaire on the Big Five traits. In reality, they might have more knowledge than what was assessed in the questionnaire.

3.4 User Behavior Based Friend Recommendation

3.4.1 System Framework Design

In this section, we will make an introduction on the ideas and frameworks of our friend recommendation system. The system is structured as follows [11]:

- Collecting user information and modeling
- User clustering
- Friend recommendation

3.4.2 User Information Collecting and Modeling

We set up a forum using discuss and embed the friend recommendation system into the forum. System records the users' behaviors (browsing and posting) in several plates and set different weights on the two kinds of records to character the users' preferences on each plate [11].

3.4.3 User Clustering

In section 3.2, we established a corresponding model for each user [11], describing the features of users on N dimensions. In the method we represented below, we use the cluster model to the system overhead when recommending potential friend. We adopt the cosine similarity to describe the degree of how much they are alike: similarity = u*v.

However, $|u|$ $|v|$ (5) where both u and v are vectors established in 3.2 as user models to depict user's interests. clustering groups people into collections, breaking a whole large package of data into several pieces.

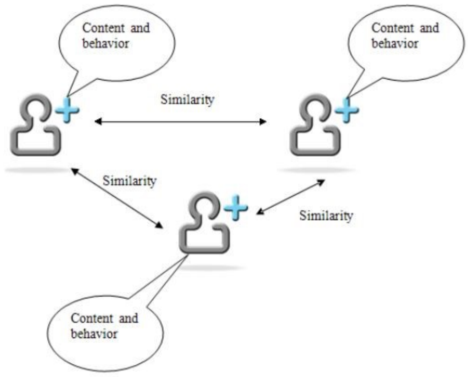


Fig. 3. Clustering users

Figure 2: Class Diagram

And in clusters we can deal with data more convenient and takes less system overload. For the reason that in each cluster the users in it are people who are most similar, the quality of recommendation is not much influenced by the reduction of users. In our system we adopted two algorithms, K-means and DBSCAN. The outcome of clustering in K-means is better. Though DBSCAN do not need the initial set of cluster number, it tends to perform badly in grouping users into several different collections or when number of dimensions goes high.

3.4.4 Implement Friend Recommendation

After the work in above, we can begin to seek the similar potential friends for users based on the clusters. In our system, we realize the recommendation in four aspects below [11]:

I. Find people who are most similar with the target user based on clusters. We calculate every people's similarity with target user and find the N-nearest ones as the raw recommendation of similar potential friends

II. Then we make the outcomes more appropriate and accurate by combining the raw potential friend's recommendation list with user's profile and the amount of their common friends, for the reason that recommendations are more persuasive coming from real friends or some common aspects (such as coming from the same school) [2]. So we adjust the outcome according to the common information between user and friend recommended

III. Another aspect that should be taken into count is that users' interests may change with time. To deal with this problem, we recalculate the clusters every one week or two weeks. The update time can be changed over different situation

IV. The new users of our system should be treated differently. They have two problems: a. new users do not have history records which can be used to create a model. b. Which cluster new users belong to is remain unknown.

This is similar to the 'cold start' problem in collaborative filtering. To deal with new ones, we use the 'friend of friend' method to recommend potential friend and adjust outcomes by getting direct feedback from users' preferences chart. When people log into a social website at the first time, they tend to add some people they already knew in the real world.

3.4.5 Works done so far

Joonhee Kwon and Sungrim Kim, proposed a method of scoring friendship based on context and friends can be recommended based on the user's physical and social information. Xing Xie, established a friend recommendation machine embedded on a bioscience web, which can recommend users who have similar research fields to each other. Jilin Chen, based on IBM's Beehive dating site compares four recommendation methods including the four friends on their web sites.

3.5 The Big Five Personality Inventory Study

The "Big Five" model of personality dimensions has emerged as one of the most well-researched and well-regarded measures of personality structure in recent years. The Big Five traits are characterized by the following:

3.5.1 Openness to Experience

Curious, intelligent, imaginative. High scorers tend to be artistic and sophisticated in taste and appreciate diverse views, ideas, and experiences.

3.5.2 Conscientiousness

Responsible, organized, persevering. Conscientious individuals are extremely reliable and tend to be high achievers, hard workers, and planners.

3.5.3 Extroversion

Outgoing, amicable, assertive. Friendly and energetic, extroverts draw inspiration from social situations.

3.5.4 Agreeableness

Cooperative, helpful, nurturing. People who score high in agreeableness are peacekeepers who are generally optimistic and trusting of others.

3.5.5 Neuroticism

Anxious, insecure, sensitive. Neurotics are moody, tense, and easily tipped into experiencing negative emotions.

4 Proposed Idea

4.1 Idea Description

Detection of human behavior can easily be done by having a close look to his/her activities in social networks. Besides the behavior, each user's personality is also analyzed. Every Social Networking platform has its own friend recommendation system which is developed using the graph based analogy. This analogy refers to for example, if I am a user, I will only get those persons in my friend recommendation list who are connected to my connections. In this paper, we will be focusing to develop an efficient and perfect recommendation system with a high accuracy. We will collect feedback from user about their behavior they will do in different situations given to them mentioned in our questionnaire. Based on that we will generate the personality profile of a user by using a standard procedure of IPIP.org. Based on these data, our system will allow the existing and/or new users to filter out what kind of persons he/she wants to get recommended. There will be specific option for selecting the personality category. which will eventually help a user to get the list of best matched users one after another. Also there will be automatically recommended friend list which will be generated based on the personality similarity between those users.

4.2 Methodology

Our core research question asks whether social media user's personality profile can be generated based on some social scenarios and based on that personality profile whether friend recommendation can be done or not. If so, then there is an opportunity to integrate the many results on the Implications of personality factors and behavior into the users' online experiences and to use social media profiles as a source of information to better understand individuals. For example, the friend suggestion system could be tailored to a user based on whether they are more introverted or extroverted.

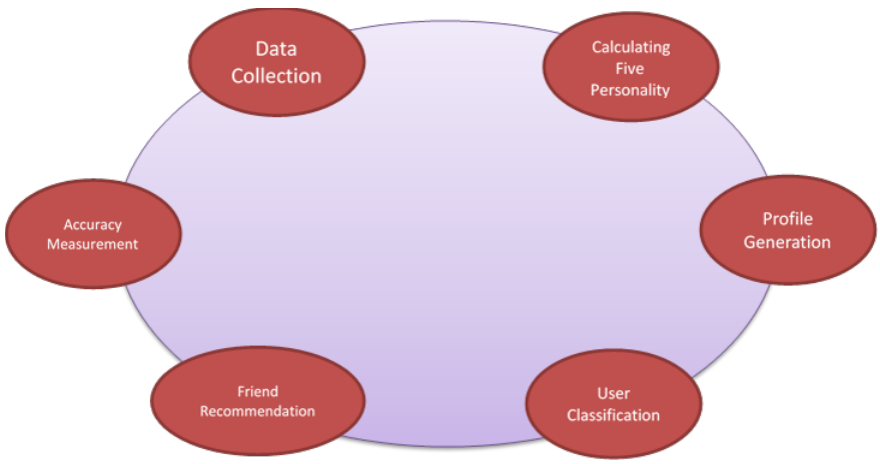


Figure 3: Our Approach

4.3 Preview of Application (*Persommend*)

In this section we compare among the trivial agglomerative method, most recently developed Improved Agglomerative Clustering Technique and our proposed method. As we can see from the table that our.

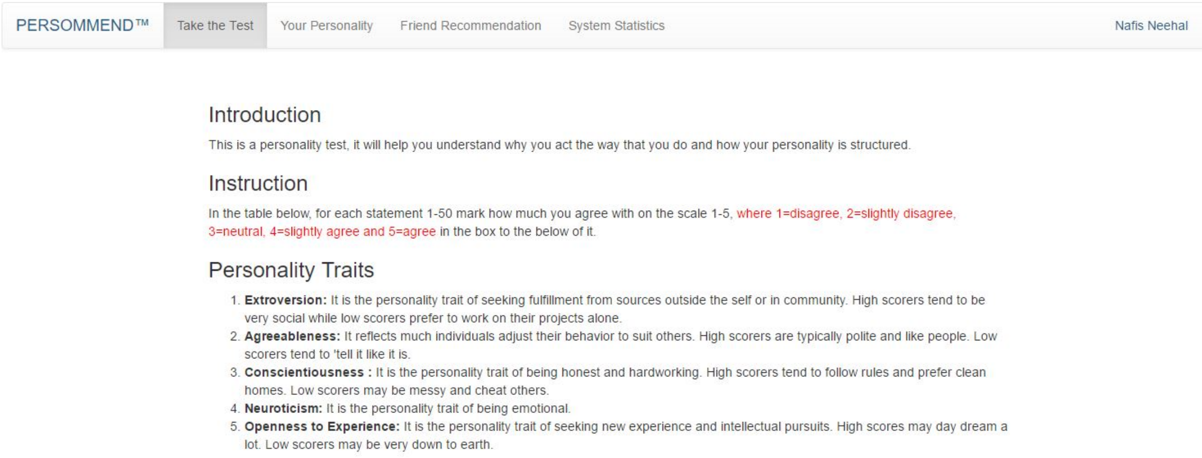


Figure 4: Persommend

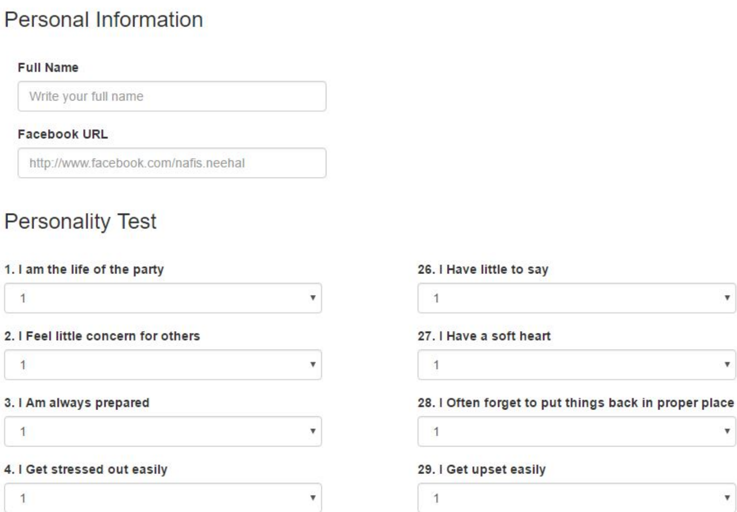


Figure 5: Persommend

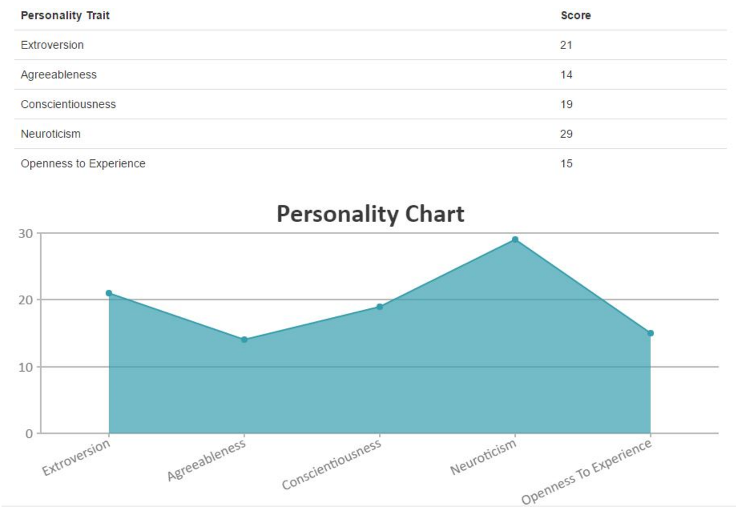


Figure 6: Persommend

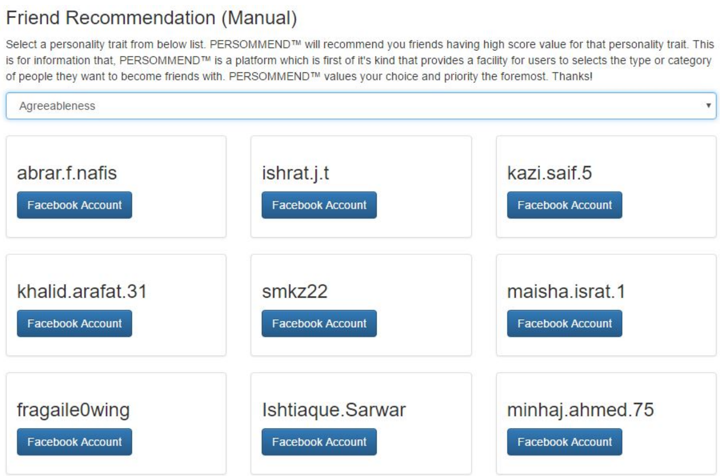


Figure 7: Persommend

5 Implementation

5.1 Data Collection

Data Collection went through different phases. We selected some arbitrary 100 facebook users. The target group was a mix of any age, profession and gender. We collected the entire data in the following way:

- We created an online survey form called "We created "Persommend" web service.
- The survey form was hosted in an online server.

- The form consists of 50 personality questions.
- Each row of the form had an option to rate the questions.
- The value that we saved from the form was numeric.
- We collected the data and kept it in to the database.
- These survey data helped us to identify big five traits of a user.

Test

Rating	I....	Rating	I.....
	1. Am the life of the party.		26. Have little to say.
	2. Feel little concern for others.		27. Have a soft heart.
	3. Am always prepared.		28. Often forget to put things back in their proper place.
	4. Get stressed out easily.		29. Get upset easily.
	5. Have a rich vocabulary.		30. Do not have a good imagination.
	6. Don't talk a lot.		31. Talk to a lot of different people at parties.
	7. Am interested in people.		32. Am not really interested in others.
	8. Leave my belongings around.		33. Like order.
	9. Am relaxed most of the time.		34. Change my mood a lot.
	10. Have difficulty understanding abstract ideas.		35. Am quick to understand things.
	11. Feel comfortable around people.		36. Don't like to draw attention to myself.
	12. Insult people.		37. Take time out for others.
	13. Pay attention to details.		38. Shirk my duties.
	14. Worry about things.		39. Have frequent mood swings.
	15. Have a vivid imagination.		40. Use difficult words.
	16. Keep in the background.		41. Don't mind being the center of attention.
	17. Sympathize with others' feelings.		42. Feel others' emotions.
	18. Make a mess of things.		43. Follow a schedule.
	19. Seldom feel blue.		44. Get irritated easily.
	20. Am not interested in abstract ideas.		45. Spend time reflecting on things.
	21. Start conversations.		46. Am quiet around strangers.
	22. Am not interested in other people's problems.		47. Make people feel at ease.
	23. Get chores done right away.		48. Am exacting in my work.
	24. Am easily disturbed.		49. Often feel blue.
	25. Have excellent ideas.		50. Am full of ideas.

Figure 8: Training Question Set

5.2 Data Preprocessing

After collecting the data, we calculated the Big five traits using certain equations predefined my ipip.org. Here goes the procedure:

- These survey data helped us to identify big five traits of a user.
- We calculated the big five traits mainly by the predefined equations.
- The equations provide a value between 40 for each traits.
- The more the value is the more that user has that personality

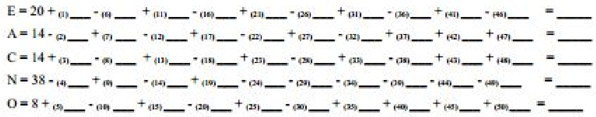
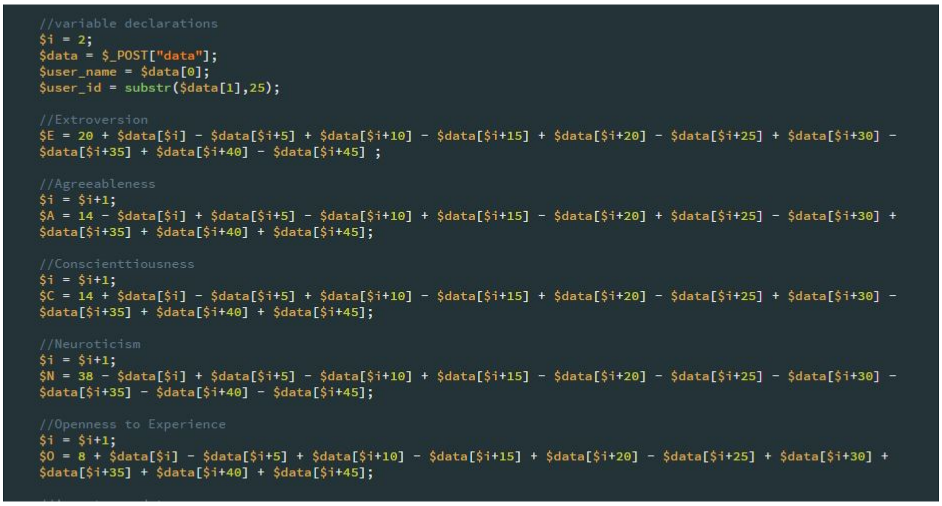


Figure 9: Equation

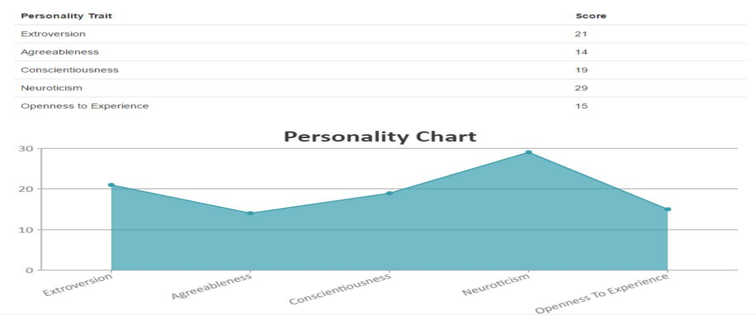


```
//variable declarations
$i = 2;
$data = $_POST["data"];
$user_name = $data[0];
$user_id = substr($data[1],25);

//Extroversion
$E = 20 + $data[$i] - $data[$i+5] + $data[$i+10] - $data[$i+15] + $data[$i+20] - $data[$i+25] + $data[$i+30] -
$data[$i+35] + $data[$i+40] - $data[$i+45] ;

//Agreeableness
$i = $i+1;
$A = 14 - $data[$i] + $data[$i+5] - $data[$i+10] + $data[$i+15] - $data[$i+20] + $data[$i+25] - $data[$i+30] +
$data[$i+35] + $data[$i+40] + $data[$i+45];

//Conscienttiousness
$i = $i+1;
$C = 14 + $data[$i] - $data[$i+5] + $data[$i+10] - $data[$i+15] + $data[$i+20] - $data[$i+25] + $data[$i+30] -
$data[$i+35] + $data[$i+40] + $data[$i+45];

//Neuroticism
$i = $i+1;
$N = 38 - $data[$i] + $data[$i+5] - $data[$i+10] + $data[$i+15] - $data[$i+20] - $data[$i+25] - $data[$i+30] -
$data[$i+35] - $data[$i+40] - $data[$i+45];

//Openness to Experience
$i = $i+1;
$O = 8 + $data[$i] - $data[$i+5] + $data[$i+10] - $data[$i+15] + $data[$i+20] - $data[$i+25] + $data[$i+30] +
$data[$i+35] + $data[$i+40] + $data[$i+45];
```

Figure 10: Code Segment

5.3 User Profile Generation

Each user will have their own profile in our system. Once they fulfilled the survey form, the profile will be generated. The user profile is calculated from the equations of International Personality Item Pool.



Personality Trait	Score
Extroversion	21
Agreeableness	14
Conscientiousness	19
Neuroticism	29
Openness to Experience	15

Figure 11: Personality Profile

5.4 User Classification

First 40 users' data is considered as the training data. We calculated the personality of each of the users and created a chart to find out which personality has the major user base.

Personality Trait	Total User Count
Extroversion	10
Agreeableness	9
Conscientiousness	17
Neuroticism	7
Openness to Experience	7

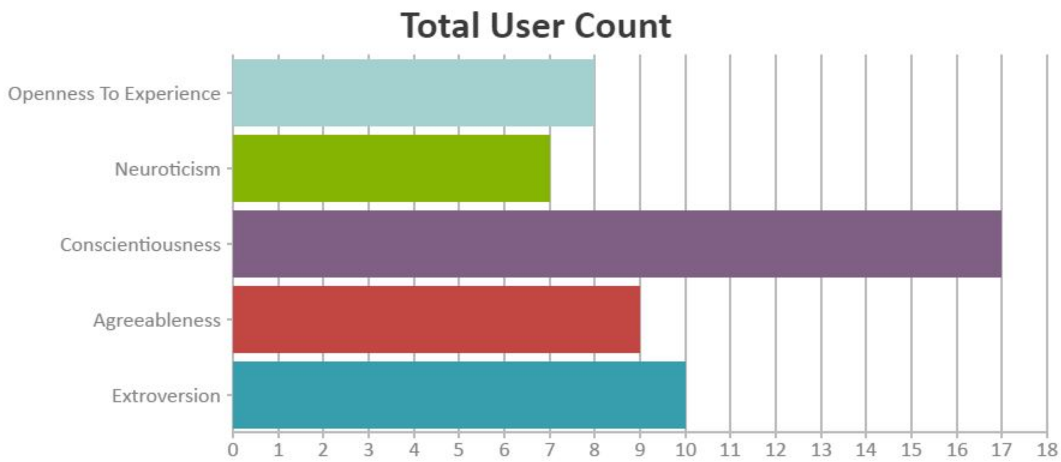


Figure 12: User Statistics

5.5 Friend Recommendation

Our system has a completely different section for friend recommendation. The recommendation system will only work within the existing users. The new user can find his/her preferred friend depending on the personality. The system will provide top five users with the choice.

5.5.1 Manual Recommendation

The users can find friends depending on their chosen personality. The users just have to select a personality trait as their preference and our system will sort out the best matched users.

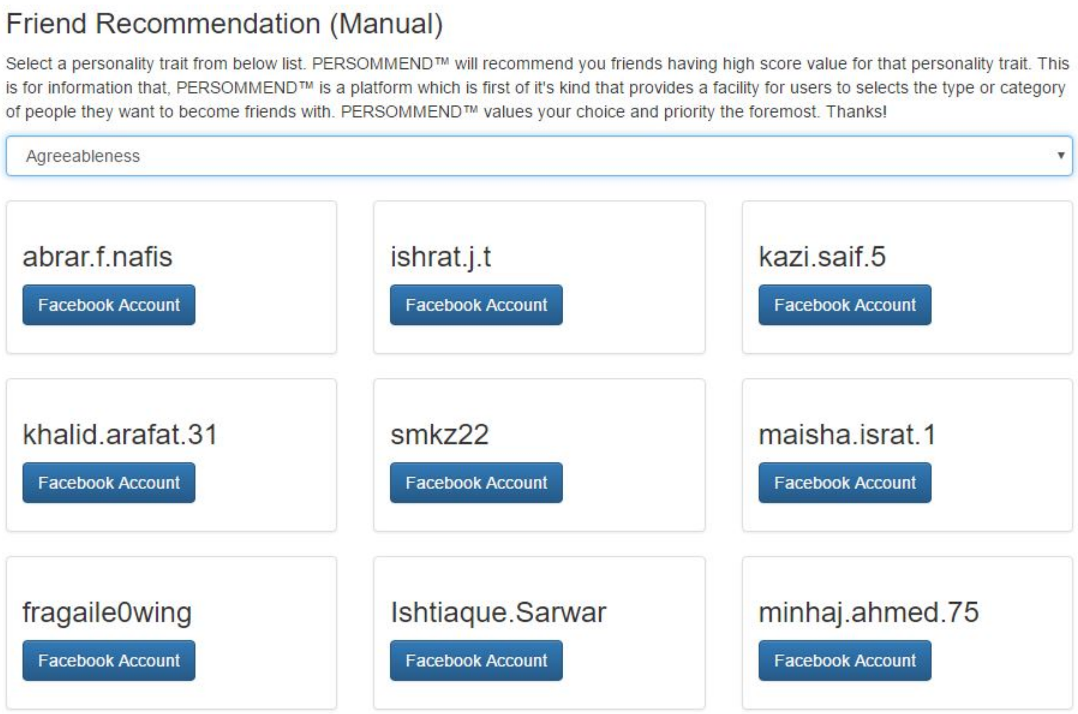


Figure 13: Friend Recommendation

5.5.2 Auto Recommendation

We implemented the cosine similarity to find out the best match for individual users. This similarity algorithm helped us to find out similar personality of a particular user. We kept the top five users from our existing user bae. This entire system is totally automatic and depends on the users personality.

$$similiarity = \cos(\theta) = \frac{A.B}{||A||||B||} = \frac{\sum_{i=1}^{n} A_{\mathrm{i}} B_{\mathrm{i}}}{\sqrt{\sum_{i=1}^{n} {A^2}_{\mathrm{i}}}\sqrt{\sum_{i=1}^{n} {B^2}_{\mathrm{i}}}} \tag{1}$$

```php
<!-- Cosine Similarity Measurement-->
<?php

    //require files
    require_once('CosineSimilarity.php');
    require_once('connect.php');

    $user_name_input = 'nafis.neehal';

    //mysql user for test data
    $queryN = "select * from user_personality where user_id='$user_name_input'";
    $resultN = mysqli_query($conn, $queryN);
    $rowN = mysqli_fetch_row($resultN);
    $user_array = array($rowN[1],$rowN[2],$rowN[3],$rowN[4],$rowN[5]);

    //mysql queryn for other user
    $query = "select * from user_personality where user_id !='$user_name_input'";
    $result = mysqli_query($conn, $query);
    $num_rows = mysqli_num_rows($result);

    //cosine similarity
    $cs = new CosineSimilarity();
    $max_cos_val = 0;
    $max_user = '';

    for($i=0; $i<$num_rows; $i++){
        $row = mysqli_fetch_row($result);
        $test_array = array($row[1], $row[2], $row[3], $row[4], $row[5]);
        $result_cosine = $cs->similarity($user_array, $test_array);
        if($result_cosine>=$max_cos_val){
            $max_cos_val = $result_cosine;
            $max_user = $row[0];
        }
        //var_dump($user_array);
        //var_dump($test_array);
    }
?>
```

Figure 14: Code Section For Cosine Similarity

```php
<?php
/**
 * CosineSimilarity measures a cosine similarity between two vectors
 *
 */

class CosineSimilarity {

  public function similarity(array $vec1, array $vec2) {
    return $this->_dotProduct($vec1, $vec2) / ($this->_absVector($vec1) * $this->_absVector($vec2));
  }

  protected function _dotProduct(array $vec1, array $vec2) {
    $result = 0;

    foreach (array_keys($vec1) as $key1) {
      foreach (array_keys($vec2) as $key2) {
        if ($key1 === $key2) $result += $vec1[$key1] * $vec2[$key2];
        }
    }

    return $result;
  }

  protected function _absVector(array $vec) {
    $result = 0;

    foreach (array_values($vec) as $value) {
      $result += $value * $value;
    }

    return sqrt($result);
  }
}
```

Figure 15: Class Section of Cosine Similarity

Cosine Similarity Matrix (50X2 dim)

```
C:\wamp64\www\persommend\recommend.php:129:
array (size=50)
  'afrin.noumin' => float 0.98618850608249
  'shuvo915' => float 0.98249293809721
  'md.hisham.56' => float 0.97101609582383
  'confused.kabir' => float 0.97087523631475
  'Asif.joy.733' => float 0.96913049635582
  'mahbubur.rahman.3323' => float 0.96589174129905
  'SaadmanHussain' => float 0.96402111207796
  'rif9201' => float 0.96126080559463
  'shuvo38' => float 0.95992721294301
  'farjana.rini.12' => float 0.95815672313206
  'Oninda' => float 0.95800604957646
  'mewreza' => float 0.95640316723279
  'rafiIUT' => float 0.94344922405498
  '1798.blackiller' => float 0.94186585684786
  'sabrina.r.oshin' => float 0.93357135045221
  'eminentkashfia' => float 0.93286628914648
  'Ishtiaque.Sarwar' => float 0.93283110496747
  'motin.nafisa' => float 0.92967593194777
  'mhshuvobd' => float 0.92868866293692
  'kadir.buet.ce' => float 0.92580318438623
  'unpredictable.yasin' => float 0.92505017450132
  'israt.nowrin' => float 0.92377723654697
  'nobody786' => float 0.91778198974294
  'jonayed.kaysar' => float 0.91215148725814
  'abrarfaiyaztasin' => float 0.91078162256946
  'salmansayeed.khan' => float 0.90214521937381
  'isovixceo' => float 0.90211226678647
  'shahrukhrydwan' => float 0.90065532654092
  'abdurrahman.maruf' => float 0.8997498887866
  'abubakar.tanvir' => float 0.89882222750723
  'mdmuntasir.therocker' => float 0.898235742119
```

Figure 16: Section of Cosine Similarity Matrix

6 Research Contribution

Our proposed system has overcome 3 of the 4 well-established problems of typical recommender system – Sparsity Problem, Over Specialization Problem and Limited Content Analysis Problem Our system facilitates user with both manual and auto mode for friend recommendation which ensures the fact that users can choose their category of friends as well as get the most relevant users recommended as friends

7 Future Works

Some of our future work plans are listed below -

- Collection of data for at least 1000 people
- Accuracy and performance measurement
- More insight provide on personality
- More customization on friend recommendation process

8 Conclusion

Recommendation, Friend Recommendation in Social Network, Personality Analysis from Social Network Data all these concepts are quite old now. Lots of works has been done in these arena. But combining these two concepts together and use this hybrid model for friend recommendation is still a new and unique concept and also a challenging and promising one. Using Collaborative Filtering Based Friend Recommendation System or the Content Based Approach has their limitations as Sparsity Problem, Cold User Problem, Over Specialization Problem, Limited Content Analysis Problem. This is because these friend recommendation systems mainly follow graph-based analysis or common user interest based method to recommend friend in social network. Analyzing user personality from social network data can be very useful in this context as per our gathered statistics. As there are already well-established methods for extracting user personality trait from his/her social network data, we can use this approach to classify users into different categories containing different personality traits and then we can recommend users belonging into the same class to each other for being friend in social network. This approach of ours can also be used in case of product recommendation, event recommendation, group recommendation and all sorts of other fields where a recommendation system is necessary.

References

[1] C. Basu, H. Hirsh, and W. Cohen, "Recommendation as classification: Using social and content-based information in recommendation," in Proceedings of the national conference on artificial intelligence. John Wiley Sons LTD, 1998, pp. 714–720.

[2] G. Adomavicius and A. Tuzhilin, "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, vol. 17, no. 6, pp. 734–749, 2005.

[3] Konstas, V. Stathopoulos, and J. M. Jose, "On social networks and collaborative recommendation," in Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval. ACM, 2009, pp. 195–202.

[4] Hotho, R. Jschke, C. Schmitz, and G. Stumme, "Information retrieval in folksonomies: Search and ranking," in The semantic web: research and applications. Springer, 2006, pp. 411–426.

[5] Song Chen, Samuel Owusu, Lina Zhou, "Social Network Based Recommendation Systems: A Short Survey" in IEEE Computer Society, 2013.

[6] Li bian, Henry Holtzman "Online Friend recommendation through Personality Matching and Collaborative Filtering" in The Fifth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, UBICOMM 2011.

[7] Namrata M. Eklaspur, Anand S. Pashupatinath, "Friend recommender system for social networks by life style extraction using probabilistic method" in International Journal of Computer Science Trends and Technology in June-2015.

[8] Lizi Zhang, Hui Fang, Wee Kiong Ng, Jie Zang "Interaction Ranking-Based trust worthy Friend Recommendation" in International Joint Conference of IEEE TrustCom-11/IEEE ICESS-11/FCST-11, in 2011.

[9] Xiwang Yang, Wang Guo, Young Liu "Bayesian-Inference-Based Recommendation in Online Social Networks", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 24, NO. 4, in APRIL 2013.

[10] Wu Youyoua, Michal Kosinskib, and David Stillwella "Computer-based personality judgments are more accurate than those made by humans" in PNAS in January 2015.

[11] Zhiwei Deng, Bowei He, Chengchi Yu, and Yuxiang Chen, "Personalized Friend Recommendation in Social Network Based on Clustering Method" in ISICA, 2012, China.

[12] Jilian Tang, Xia Hu, Huan Liu, "Social Recommendation: A Review" in 2013, China.

[13] Vinti Agarwal, K. K. Varadwaj "A collaborative filtering framework for friend recommendation in social networks based on interaction intensity and adaptive user similarity." In September, 2012.

[14] Jennifer Gobeck, Christina Robles, Karen Turner "Predicting Personality with Social Media" in 2011

[15] IPIP.ori.org standard questionnaire on Big Five Personality Analysis "The Big Five Personality Test"