Volume 7, Issue 5, May 2017 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Solving Cold-Start Problem by Combining Personality Traits and Demographic Attributes in a User Based Recommender System

Gizaw, Zewengel Tilahun*, Huang Dong Jun, Ammar Oad
Department of Computer Science, Central south University, Changsha,
Hunan, China

***Abstract-** Several approaches have been suggested for providing users with recommendations using their rating history, most of these approaches suffer from new user problem (cold-start) which is the initial lack of items rating. Most hybrid approaches use the combination of the rating based similarity measure with demographic filtering to solve the cold-start problem. In this paper we combined the rating based similarity calculation with the personality traits based similarity and demographic filtering. We integrated personality traits with the demographic characteristics because users with the same demographic attributes may have different personality or behavior so personality information is a necessary filtering condition for users to find better reference users to be recommendation guides. Even though users don't have a rating history the proposed enhanced similarity measure will find users with the same demographic attributes like age, gender, location…etc and the same personality traits. This helps to find the k-nearest neighbors for new users and solve cold-start problem which also can improve the performance of the recommender system.*

***Keywords:** recommender system, cold-start problem, demographic filtering, personality traits, similarity measures.*

I. INTRODUCTION

With the beginning of the Web 2.0 era, the internet began growing up and developing with tremendous speed. Many opportunities, such as sharing knowledge, information, opinion today every user of the World Wide Web can purchase almost any item being in any country of the world. As opposed to real shops, in the internet there are no place limitations. In fact, there is almost endless place. Nevertheless people came across a new problem in the WWW. The amount of information and items got extremely huge, leading to an information overload. It became a big problem to find what the user is actually looking for. Search engines partially solved that problem, however personalization of information was not given. So developers found a solution in recommender systems. Recommender systems are tools for filtering and sorting items and information. They use opinions of a community of users to help individuals in that community to more effectively identify content of interest from a potentially overwhelming set of choices. There is a huge diversity of algorithms and approaches that help creating personalized recommendations [3].

The internal functions for RS are characterized by the filtering algorithm. The most widely used classification divides the filtering algorithms into:

(1) Collaborative filtering
(2) Demographic filtering
(3) Content-based filtering
(4) Hybrid filtering

The fundamental assumption of CF is that if users X and Y rate n items similarly, or have similar behaviors (e.g., buying, watching, listening), and hence will rate or act on other items similarly. CF techniques use a database of preferences for items by users to predict additional topics or products a new user might like. In a typical CF scenario, there is a list of m users {u1, u2. . . um} and a list of n items {i1, i2, . . . , in}, and each user, ui, has a list of items, Iui, which the user has rated, or about which their preferences have been inferred through their behaviors. The ratings can either be explicit indications, and so forth, on a 1–5 scale, or implicit indications, such as purchases or click-troughs.

User-based Collaborative Filtering has been studied in-depth during the last decades. It is one of the most successful and widely used recommendation technologies, owing to its compelling simplicity and excellent quality of recommendations. It assumes that if a group of users have similar interests in their previous behaviors, they will express similar interests on other more items in the future. Its basic idea is to find a group of users, who have a History of agreeing with an active user (i.e., they either gave similar ratings or purchased similar items). Once a neighborhood of users is formed, opinions from these neighbors are aggregated to produce recommendations for the active user [12].

User-based algorithms, also known as neighborhood-based, are one of the most popular strategies of collaborative filtering. They follow a three-step process.

(1) Calculate the similarity between the active user and the rest of the users.
(2) Select a subset of the users (neighborhood) according to their similarity with the active user.
(3) compute the prediction using the neighbor ratings [6], [20].

Collaborative systems have their own limitations, as described below.

- **Data Sparsity** problem arises if the user-item matrix containing ratings details is extremely sparse and this situation further leads to inefficient recommender systems which are based upon **nearest-neighbor algorithms** for calculating user similarity. This problem is further classified as **reduced coverage problem and neighbor transitivity problem.**
- **Scalability** problem refers to a situation when numbers of items and users giving ratings those items increased tremendously and it becomes difficult for a recommender system to handle such a big data due to computational complexity and constrained resources. So, it goes beyond the limit of acceptability of recommender systems.
- **Synonymy** means recommender system fails to recognize the similarity among two items when some similar items have different names and the recommender system treats them as if they are different items. This leads to problem of recommending similar items, called as synonymy problem.
- **Gray sheep** problem arises because the user's choice does not match with any other user or group of users in agreement or disagreement consistently.
- **Black sheep** problem refers to a situation when the user's choice correlates with very few users or no users at all. In such cases recommender system proves to be inefficient or fruitless in generating preferences.
- **Shilling attacks** are categorized into **push attacks and nuke attacks.** When the competitor vender makes use of unfair ways and means to show more rating of their own items as compared to other vendor products then it is known as push attacks. On the other hand if they try to reduce the rating of their rivals or competitors then it is called as nuke attacks.
- **Cold-start problem** means when a recommender system is unable to generate or predict ratings due to initial lack of ratings then it is referred to as cold-start problem. This kind of situation happens when either a new user arrives into a system having no rating records available with recommender system or when a new item enters into system and till now no one has given rating to that item. So, it becomes difficult for a recommender system to generate choices for a new user and hence the objective of recommender systems is not achieved.

This paper can be outlined as follows. Section 2 provides the literature review. Section 3 methodology and proposed system architecture. Section 4 presents an overview of similarity measures involved in our experiments. Section 5 describes our experimental work and evaluation. Section 6 concludes the paper and provides directions for the future.

II. LITERATURE REVIEW

In this Section we review the main body of existing work relevant to my proposed research. The new user problem makes it hard for the system to learn about the new user preferences especially if he/she did not rate enough items; this is also known as cold-start users. The new item problem is also an issue for collaborative filtering since even if the item has a high rating, the recommender system will not be able to recommend it unless a minimum number of users have rated it. Sparsity of data which means that the number of existing ratings is relatively low compared to the number of users and items on the system. This can impact negatively the accuracy of the recommender system. Due to the above limitations, hybrid models were developed. Badaro, H. Hajj, W. El-hajj, and L. Nachman [4] propose a hybrid model that combines simultaneously user-based collaborative filtering and item-based collaborative filtering by adding the predicted ratings from each technique and multiplying them with a weight that incorporates the accuracy of each technique alone. S. Panigrahi, R. K. Lenka, and A. Stitipragyan[14] presents a new hybrid solution to user based traditional CF methods based on the Apache Spark platform combining both dimension reductionality and clustering methods of machine learning S. Xia, S. Chen, and Z. Wang [24] proposed a new similarity function considering the weight of item. He start from two aspects of item. At first, the similarity of target item with the other items is attached into the traditional algorithms, which leads to more accurate neighbors for every item of target user. B. Chikhaoui, M. Chiazzaro, and S. Wang [8] introduced an efficient hybrid approach for recommender systems. Three recommendation algorithms are incorporated in their approach to predict the ratings. Other researchers modified user similarity calculation method to employ the hybridization of demographic and collaborative approaches. A modification to k-nearest neighborhood had been introduced which calculates the similarity scores between the target user and other users forming a neighborhood, increasing the scores of users having similar ratings and demographic attribute[28]. Whereas another research work demonstrated another modified version of k-nearest neighborhood by adding a user demographic vector to the user profile, the similarity calculation consider both ratings and demographic vector [22].

III. METHODS AND PROPOSED SYSTEM ARCHITECTURE

The central contribution of this paper to the research community is to solve the cold-start problem by combining user's personality traits and demographic attributes in order to get the k-nearest neighbors for new user which leads us to find the accurate similar users even though the new user does not have rating history. The most important task in user based recommender system is finding the similar users using the appropriate similarity measure because different measurements lead to different neighbor users, in turn, leading to different recommendations.

In this research we proposed the combination of three types of similarity calculations by doing this we can come up with a better user's similarity measure for user based recommender system to get the k-nearest neighbors for new user. To come with the hybrid recommender system, mixed research methods approach will be used. We intend to start with a thorough examination of both the published and unpublished material related to pure recommender systems that implement user based collaborative filtering, demographic filtering and the hybrid recommender system. we combined the two most important information about user because we assume that people with the same demographic attributes may have different personality or behavior so our similarity measure will perform the combination of users with similar demographic attributes and personality traits so as to get the better k-nearest neighbors for new user that does not have rating history, because the most important step in a user based recommender system is finding the similarity of users. So we used three similarity calculations, similarity based on user's interest change using the improved Pearson correlation coefficient, similarity based on user's personality and similarity based on user's demographic attributes.

The combination of approaches can proceed in different ways:

1. Separate implementation of algorithms and joining the results.
2. Utilize some rules of user-based collaborative filtering in demographic and personality approaches.
3. Create a unified recommender system that brings together the whole approaches.

As mentioned earlier the basic idea of this framework is considering the combination of users rating, user's personality traits and their demographic attributes. First User's similarity will be calculated based on their rating using the improved Pearson correlation coefficient if the user is a new user the cold-start problem take place so the system will solve this problem by calculating users similarity based on their demographic attributes and personality traits.

Our research objective is not to build k nearest neighbors rather to proof that cold-start problem will be solved in a better way and when the similar attributes of two users increase their correlation also increase that is why we mixed users demographic attributes with their personality traits, trying to solve cold-start problem using only demographic characteristics cannot be a complete solution because users with the same demographic characteristics may have different personality or behavior.

Most researches use demographic filtering to solve cold-start problem if users do not have rating history the similarity among them will be calculated based on their demographic attributes, this means if u1 is similar with u2 according to age, gender and so on this implies u2 can recommend to u1 and vice versa. The problem happens when two users in the similar age level, gender and with the same occupation may have different personality traits. Because Users interest differ based on their personality. [9], [10], [13].

So the recommendation based on only demographic characteristics may drive us to the wrong decision by recommending items to the mismatch users. If we add the demographic attributes and the personality traits of users and calculate their similarity the result will be better and accurate even though the two users do not have rating history.

A. Benchmark Methods

The methods in comparisons with our proposed one are as follows:

1. *Similarity Based on user interest change*: This is a method which calculates the correlation between two users based on their rating that they give to similar items and consider the time the user rated the item as one factor to determine the similarity this means, the rating a user give to an item now may be changed through time because of users interest change this means rating is not a constant value. Even though cold-start problem is still there this method is better than the traditional Pearson correlation coefficient. The output will be -1 which is the lowest and 1which is the highest possible value.

   i. Pearson similarity calculation of two users' u and v [5].

$$sim(u,v) = \frac{\sum_{i=1}^{m}(R_{u,i}-\overline{R_u})(R_{v,i}-\overline{R_v})}{\sqrt{\sum_{i=1}^{m}(R_{u,i}-\overline{R_u})^2}\sqrt{\sum_{i=1}^{m}(R_{v,i}-\overline{R_v})^2}} \quad (1)$$

Where $R_{u,i}$ and $R_{v,i}$ are the rating of user u and v for item i , and , $\overline{R_u}$ and $\overline{R_v}$ are their respective average rating. m is the number of items.

   ii. Similarity based on user interest change[24]

$$sim(u,v) = \frac{\sum_{j=1}^{m} Isim(i,j)^2 * w(u,j) * (R_{uj} - \overline{R_u})}{\sqrt{\sum_{j=1}^{m}(Isim(i,j) * w(u,j) * (R_{ui} - \overline{R_u}))^2}} * \frac{\sum_{j=1}^{m} w(v,j) * (R_{vj} - \overline{R_v})}{\sqrt{\sum_{j=1}^{m}(Isim(i,j) * w(v,j) * (R_{vi} - \overline{R_v}))^2}} \quad (2)$$

Sim (u, v) is the similarity between users u and v, Isim (I, j) the similarity between item I and j we use the Pearson as similarity function, w (u, j), w (v, j) is the weight of time.

The closer the time is to now, the more likely it is to reflect the actual situation.

$$W(u,i) = (1-\alpha) - \alpha\frac{T_{ui}}{T_u} \quad (3)$$

α is a factor to control the weight of time.$\alpha \epsilon$ (0,1) . If the $\alpha$is greater, the weight of time will be larger in the similarity calculation.

2. *Similarity Based on users Demographic attributes:* this method considers only the demographic attributes of users and most researches used this method to alleviate the cold-start problem [22].

$$Dem\_sim(u,v) = \cos\_sim(\vec{u},\vec{v})\ or\ vect\_sim(\vec{u},\vec{v}) = \frac{\vec{u}.\vec{v}}{||\vec{u}||*||\vec{v}||} = \frac{\sum_{i=1}^{m} u_i v_{,i}}{\sqrt{\sum_{i=1}^{m} ui^2}\sqrt{\sum_{i=1}^{m} vi^2}} \quad (4)$$

Dem_sim is the demographic similarity, cos_sim is the cosine similarity, vect_sim is the vector similarity. Where u_i and v_i are components of vector u and v respectively.

Table1.Example of Users Demographic Data

Name	Age	Gender	Occupation	Income	Location/Country
Mark	25	M	Teacher	10000 ¥	China
Alice	19	F	Student	3000¥	France
Paul	60	M	Engineer	15000¥	England
Mary	35	F	Nurse	5000¥	Russia
Mike	20	M	Student	2500¥	South Africa

3. *Similarity Based on user personality:* here the personality similarity between two users' u and v can be computed as the Pearson correlation coefficient of their personality descriptors [9], [10], [13].

Table2.Users Personality descriptor matrix.

u/p	E	A	C	ES	O
U1	4.56	5.26	5.47	4.85	5.43
U2	4.43	5.21	5.45	4.90	5.53
U3	4.12	5.14	5.11	4.64	5.07
U4	4.38	5.37	5.57	5.14	5.53
U5	4.07	5.21	5.34	4.89	5.43

E: Extroversion, A: Agreeableness, C: Conscientiousness, ES: Emotional Stability, O: Openness

$$simp(u,v) = \frac{\sum_k (P_u - \overline{P_u})(P_v - \overline{P_v})}{\sqrt{\sum_k (P_{u,} - \overline{P_u})^2}\sqrt{\sum_k (P_{v,} - \overline{P_v})^2}} \quad (5)$$

Where simp (u, v) is the similarity based on personality of the users u and v and $\overline{P_u}$ and $\overline{P_v}$ are the average personality of the users u and v.

B. The Extended Demographic similarity

Here we will build a new users demographic and personality attributes table (Table 3) that incorporates the user's personality traits as one of the demographics attributes of users by adding the demographic similarity and the personality based similarity using the formula below.

Edem_sim (u, v) =Dem_sim (u, v) + simp (u, v) (6)

$$\text{Edem\_sim}(u, v) = \frac{\sum_{i=1}^{m} u_i v_{,i}}{\sqrt{\sum_{i=1}^{m} \mathrm{ui}^2}\sqrt{\sum_{i=1}^{m} \mathrm{vi}^2}} + \frac{\sum_k (P_u - \overline{P_u})(P_v - \overline{P_v})}{\sqrt{\sum_k (P_{u,} - \overline{P_u})^2}\sqrt{\sum_k (P_{v,} - \overline{P_v})^2}}$$

Where Edem_sim (u, v) is the extended demographic similarity between the user's u and v, Dem_sim (u, v) is the demographic similarity of user's u and v and simp (u, v) is the similarity based on personality of user u and v.

C. The proposed similarity measure (Enhanced Similarity calculation)

Here we proposed a similarity calculation that simply multiply the similarity based on users rating and the similarity based on the extended demographic attributes. The enhanced similarity was defined as follows:

Enh_sim (u, v) = α*simr (u, v) + (1- α)* Edem_sim (u, v) (7)

Where Enh_sim (u, v) is the enhanced similarity between the users' u and v, simr (u, v) is the rating based similarity between the users' u and v, Edem_sim (u, v) is the similarity based on the extended demographic values of user's u and v. α is a weight parameter which controls the percentage.

If users do not have ratings their similarity will be calculated based on their personality traits and demographic characteristics.

D. The demographic and personality based recommendation process

The demographic and personality based recommendation process performs four stages: data input, similarity calculation, neighborhood formation and recommendation calculation (as shown in fig.1). Data input is the stage which holds new target users demographic and personality data (the user who requires recommendations) and also ratings, demographic and personality data of the rest of users. User should explicitly determine his/her demographic information and answer the personality quizzes TIPI (Ten Item Personality Inventory) to assess users personalities (Big Five Personality Traits) and users separately took personality test to determine Big 5 personality score (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism) [9], [10], [13]. Similarity calculation stage utilizes users demographic and personality data to obtain a number of users having similar demography and personality to the target user forming a neighborhood. Finally, recommendation calculation stage obtains items which have been commonly positive-rated by neighborhood users to be suggested to the target user. It assumes that users with similar demographic attributes and personality traits will rate items similarly.

Furthermore the similarity calculation stage requires selecting the demographic and personality attributes to be used for calculating the similarities. For example, Table 3 demonstrates the demographic and personality of five users; each user has five demographic attributes and a personality descriptor. Let us assume that Mark is a new user who demand recommendation, the system has to calculate the similarity between Mark and other users based on the selected demographic attributes and the personality of Mark and other users. The similarity calculation output will be the sum of demographic similarity and the personality similarity and depends on the way the system interprets how users are similar based on the demographic and personality attributes; if users having the same gender are similar then Mark, Paul and Mike are similar, if users having the same occupation are similar then Alice is Similar to Mike. Therefore, the choice of demographic attributes affects the similarity calculation output which consequently influences the results of recommendation calculation. If we take personality as the measurement of similarity, personality may differ according to culture, religion and location. For example let's take two girls with the same age one in Beijing and the other in Urumqi they may not have the same liking for the same item because of the culture, religion and the difference in the development of the cities (two persons one in the city and the other in the countryside with the same age and gender attributes may differ in their personality). So we cannot recommend them the same item because they are similar in gender and age we should have their personality information.

E. Motivation

- Standard Matrix Factorization ignores external information about users and items.
- By analyzing users' written reviews, we have a better understanding of users.
- This understanding leads to better recommendations
- Even if users give same numeric rating for item, an analysis of the written review helps to understand how likings differ in the future.

Table 3. Example of Users Extended Demographic Data

Name	Age	Gender	Occupation	Income	Location/Country	Personality Descriptor
Mark	25	M	Teacher	10000 ¥	China	E
Alice	19	F	Student	3000¥	France	A
Paul	60	M	Engineer	15000¥	England	C
Mary	35	F	Nurse	5000¥	Russia	ES
Mike	20	M	Student	2500¥	South Africa	O

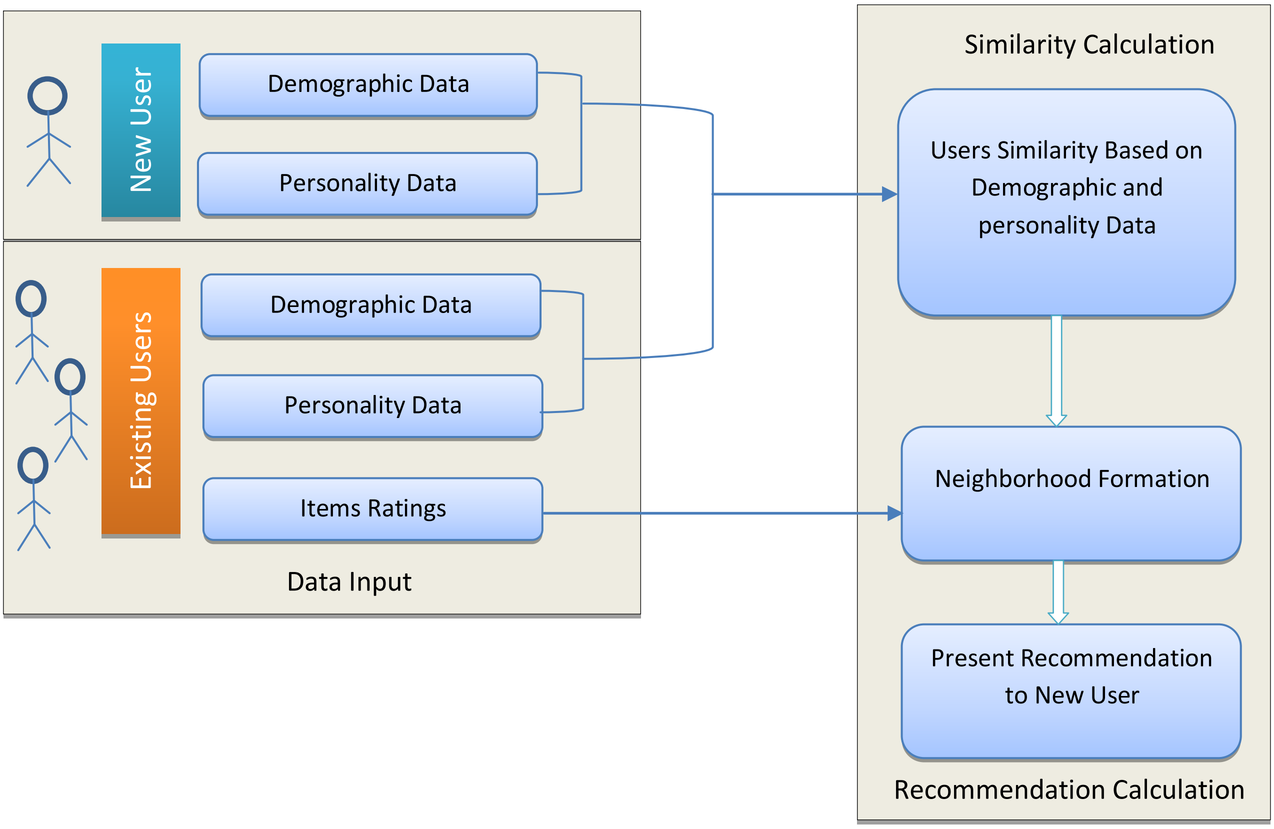

Fig 1. Demographic and Personality Based approach for new users.

F. The overall system architecture and the steps

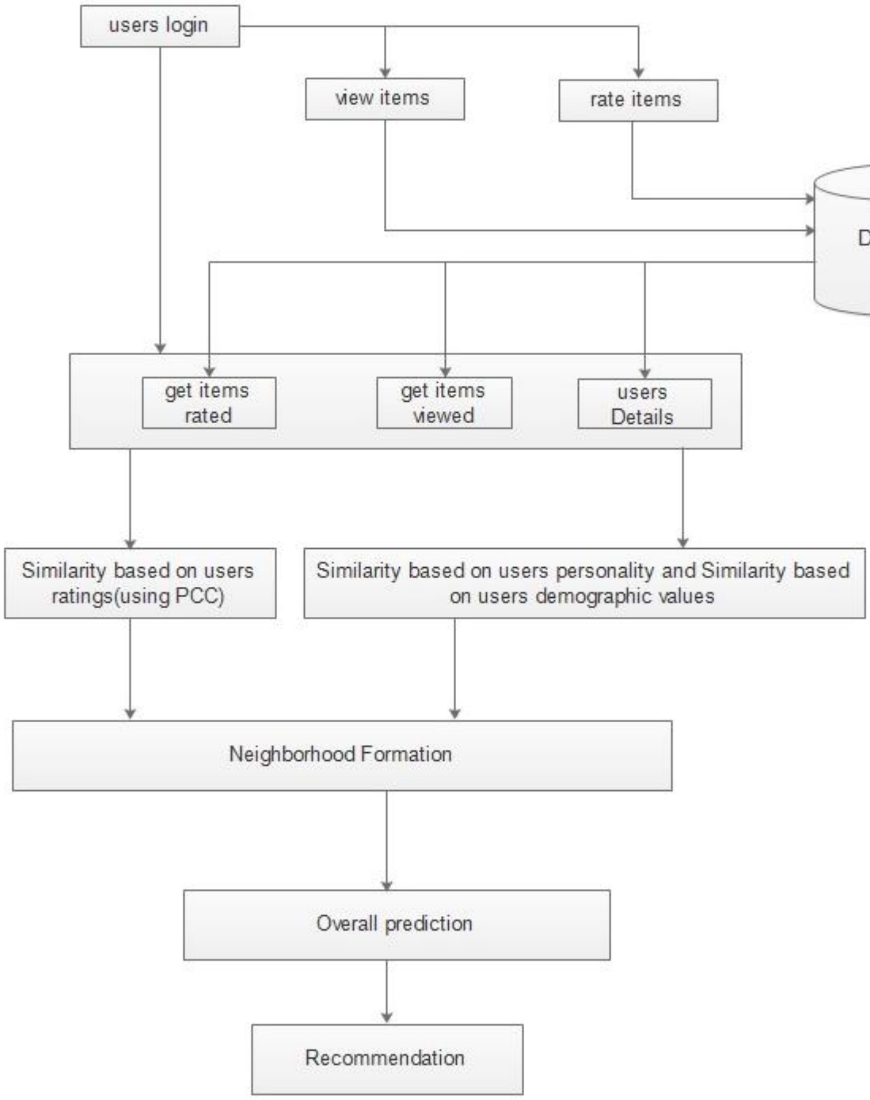

Fig 2. Shows the separate process of similarity measures, combined neighborhood formation and prediction.

The proposed framework contains the following steps:-

1. Reading user profile from the data base and keep the user profile in the main memory (for memory based approach).
   Two types of user profile:
   - Demographic Attributes
   - Personality Traits
2. Applying similarity measure based on the rating history
3. Applying similarity based on the demography and personality.
4. Then neighborhood formation.
5. Selecting items for providing recommendation to the active user.
6. Showing the recommended items to the active user.

IV. PREDICTION AND RECOMMENDATION

To obtain predictions or recommendations is the most important step in a collaborative filtering system. In the user-based CF algorithm, a subset of nearest neighbors of the target user are chosen based on their similarity with him or her, and a weighted aggregate of their ratings is used to generate predictions for the target user. In this paper, we use the modified method of plain user based prediction method to predict the rating of a target user.as shown below, for that purpose, the formula used in plain User-based Filtering [4],[6] was modified as follows:

$$Pu,i = \overline{R_u} + \frac{\sum_{v=1}^{l} (R_{v,i}-\overline{R_v})*Enh\_sim(u,v)}{\sum_{v=1}^{l} |Enh\_sim(u,v)|} \quad (8)$$

P_{ui} is prediction for the active user u on item i, $R_{v,i}$ is the rating of existing user v on item i, $\overline{R_u}$ and $\overline{R_v}$ are the average rating of the existing user u and v, l is the number of users and enh_sim u,v is the enhanced similarity.

The only difference from prediction generation, as executed in plain User- based Collaborative Filtering, being the enhanced similarity factor.

V. EXPERIMENT AND EVALUATION

Our research objective is to improve the performance of user based recommender system by designing a new hybrid recommender architecture in order to solve the cold-start problem by combining user's similarity based on their personality with demographic filtering. In the previous sections we reviewed a number of research areas which we will draw upon throughout this project. The core aspects of our project can be divided into four distinct phases which broadly encapsulate the processes of data collection, training and testing, prediction and recommendation then performance evaluation. We conduct the experimental evaluation of the proposed approach that consider users interest change, user's personality and demographic attributes then evaluate the performance of the prediction. We use the traditional collaborative filtering approach as the performance benchmark.

A. Datasets

To evaluate our method, we used experimental data from MovieLens dataset. This data set consists of 1000,209 ratings (1-5) from 6040 users on 3900 movies and the demographic attributes age, gender, occupation. Each user has rated at least 20 movies. The sparsity level is 1-1000209(3900*6040) = 0.958, the dataset is highly sparse. We randomly picked 70% data for training and the rest for testing. And a music dataset that have been accumulated in the previous study of users personality for recommender system [26] will be used, here users were asked to answer a personality quizzes based on the big five model to build users profile[9],[10]. By using the above two datasets, we'll compare our proposed method with the other benchmark methods. 20% of the ratings from the dataset will be extracted randomly to act as the test set, with the remaining be the training set.

B. Performance Evaluation

The quality of a recommender system can be decided on the result of evaluation. The type of metrics used depends on the type of CF applications. In this paper, we use Mean Absolute Error (MAE) as the measure for performance evaluation. The MAE measures the difference, as absolute value, between the prediction of the algorithm and the real rating. Despite its limitations when evaluating systems focused on recommending a certain number of items, the simplicity of its calculation and its statistical properties have made it become one of the most popular metrics when evaluating recommender systems. It is computed over all the ratings available in the evaluation subset [1], [5], [20].

$$MAE = \frac{\sum_{i=1} |P_{u,i} - R_{u,i}|}{N} \quad (9)$$

N is the total number of items in the test set, p_{ui} the predicted rating for user u on item I, R_{ui} the actual rating for user u on item i. so the smaller value of MAE means the higher quality of recommender or the better the prediction or the more accuracy the recommendation is provided by algorithm.

C. Evaluation Result

After performing the experimental settings based on the selected dataset, the results obtained from evaluating the accuracy of our proposed method compared with the rest of the methods on the benchmark. The significant differences of MAE may clearly express our enhanced Similarity method can make a more accurate and high-quality prediction and recommendation than the rest of similarity calculation methods.

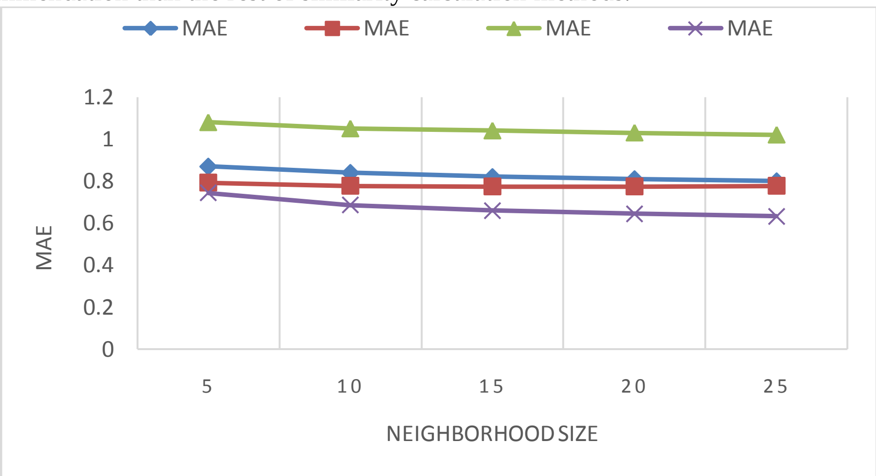

Fig 3.comparing the performance of the proposed method with the rest on the benchmark

Simr: similarity based on user interest change. *Dem_sim*: similarity based on user demographic value. *Simp*: similarity based on user personality. *Edem_sim: Extended demographic similarity*. *Enh_sim*: the enhanced similarity

The result in Fig.3. Shows that the hybrid similarity can achieve the best recommendation quality on MAE. In MAE, the smaller value is better, as we can see on the graph the proposed method (Enh_sim) value of MAE is getting less and less as the number of neighbors increased.

The MAE variation for each method in different size of neighbors, indicating that the MAE of our method is smaller than that of the other three methods at each neighbor size. The significant differences of MAE may clearly express our method can make a more accurate and high-quality prediction and recommendation than the others.

VI. CONCLUSION

With the development of society, recommendation systems are increasingly being used widely. using the traditional similarity measure alone that is based on only users rating cannot calculate the similarity of new users with the existing users because they do not have rating history so cold-start problem will happen. The main goal of this paper is to overcome cold-start problem by combining user's demographic attributes with their personality traits. So In this paper, we proposed the product of similarity functions based on users rating and the enhanced users similarity function that we got by combining users demographic attributes and personality traits which helps to find the k-nearest neighbors for those users who do not have rating history. This approach will solve the cold-start problem as well as improve the similarity measure and the performance of recommendation system apparently.

In this work we have presented a unique filtering approach that draws ideas from existing algorithms and combines them with demographic and personality information. Enh_sim is the hybrid algorithms that resulted by combining demographic filtering and personality traits based filtering to enhance user-based collaborative filtering. The evaluation result showed that our approach can perform better depending on neighborhood size. The demographic information existing in the MovieLens data set do not hold enough data about the user in order to capture their distinguishing features and generate accurate predictions, when utilized just by themselves. That is why we incorporate the income level and location as a demographic attributes which is not available in the MovieLens data set but if we use them it would be better to distinguish the users more. Still, when combined appropriately with other forms of filtering, such as personality based filtering, they can enhance the recommendation process.

In the future work, we would like to perform this experiment with more accuracy and considering that the system can get users demographic and personality data implicitly and better suit mobile users when a user changes geographical location his demographic values should be updated dynamically without any user involvement. Thus, the current recommendation system needs improvement for present and future requirements of better recommendation qualities.

ACKNOWLEDGEMENT

This work was supported by the project of National Science Fund of China(NO.60973188).

REFERENCES

[1] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems : A Survey of the State-of-the-Art and Possible Extensions," vol. 17, no. 6, pp. 734–749, 2005.

[2] H. J. Ahn, "A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem," Inf. Sci. (Ny)., vol. 178, no. 1, pp. 37–51, 2008.

[3] D. Asanov, "Algorithms and Methods in Recommender Systems," Other Conf., 2011.

[4] G. Badaro, H. Hajj, W. El-hajj, and L. Nachman, "A Hybrid Approach with Collaborative Filtering for Recommender Systems."

[5] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," Knowledge-Based Syst., vol. 46, pp. 109–132, 2013.

[6] F. Cacheda, V. Carneiro, D. Fernández, and V. Formoso, "Comparison of collaborative filtering algorithms," ACM Trans. Web, vol. 5, no. 1, pp. 1–33, 2011.

[7] T.-H. . Cheng, H.-C. . Chen, W.-B. . Lin, and Y.-H. . Lee, "Collaborative filtering with user interest evolution," PACIS 2011 - 15th Pacific Asia Conf. Inf. Syst. Qual. Res. Pacific, 2011.

[8] B. Chikhaoui, M. Chiazzaro, and S. Wang, "An improved hybrid recommender system by combining predictions," Proc. - 25th IEEE Int. Conf. Adv. Inf. Netw. Appl. Work. WAINA 2011, no. March, pp. 644–649, 2011.

[9] S. D. Gosling, P. J. Rentfrow, and W. B. Swann, "A very brief measure of the Big-Five personality domains," J. Res. Pers., vol. 37, no. 6, pp. 504–528, 2003.

[10] R. Hu and P. Pu, "Using Personality Information in Collaborative Filtering for New Users," pp. 17–24.

[11] C. M. Lewandowski, N. Co-investigator, and C. M. Lewandowski, Machine Learning with Spark, vol. 1. 2015.

[12] M. Madhukar, "Challenges & Limitation in Recommender Systems," vol. 4, no. 3, pp. 138–142.

[13] M. A. S. N. Nunes, S. A. Cerri, and N. Blanc, "Improving Recommendations by Using Personality Traits in User Profiles," J. Pers., vol. 32, no. 5, pp. 92–100, 2008.

[14] S. Panigrahi, R. K. Lenka, and A. Stitipragyan, "A Hybrid Distributed Collaborative Filtering Recommender Engine Using Apache Spark," Procedia Comput. Sci., vol. 83, no. BigD2M, pp. 1000–1006, 2016.

[15] Z. Quan, "Collaborative filtering recommendation based on user personality," pp. 307–310, 2013.

[16] A. Rajaraman and J. D. Ullman, "Mining of Massive Datasets," Lect. Notes Stanford CS345A Web Min., vol. 67, p. 328, 2011.

[17] L. Safoury and A. Salah, "Exploiting User Demographic Attributes for Solving Cold-Start Problem in Recommender System," Lect. Notes Softw. Eng., vol. 1, no. 3, pp. 303–307, 2013.

[18] S. Saini and L. Banda, "I MPROVING S CALABILITY I SSUES USING GIM IN," vol. 4, no. 1, pp. 600–610, 2012.

[19] S. Solanki, "Recommender System using Collaborative Filtering and Demographic Characteristics of Users."

[20] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," Adv. Artif. Intell., vol. 2009, no. Section 3, p. 4, 2009.

[21] P. Symeonidis, A. Nanopoulos, A. N. Papadopoulos, and Y. Manolopoulos, "Collaborative recommender systems: Combining effectiveness and efficiency," Expert Syst. Appl., vol. 34, no. 4, pp. 2995–3013, 2008.

[22] M. Vozalis and K. Margaritis, "Collaborative filtering enhanced by demographic correlation," AI Symp. Prof. Pract. AI, 18th World Comput. Congr., pp. 1–10, 2004.

[23] U. H. Wanaskar, S. R. Vij, and D. Mukhopadhyay, "A Hybrid Web Recommendation System Based on the Improved Association Rule Mining Algorithm," J. Softw. Eng. Appl., vol. 2013, no. August, pp. 396–404, 2013.

[24] S. Xia, S. Chen, and Z. Wang, "An Hybrid Similarity Function for Neighbor Selection in Collaborative Filtering," vol. 8, no. 6, pp. 243–252, 2015.

[25] L. Zhang, Q. Tao, and P. Teng, "An Improved Collaborative Filtering Algorithm Based on User Interest," J. Softw., vol. 9, no. Iccmcee, pp. 999–1006, 2014.

[26] R. Hu and P. Pu, "A Study on User Perception of Personality-Based Recommender Systems," Proc. 18th Int. Conf. User Model. Adapt. Pers. (UMAP 2010), vol. 6075, pp. 291–302, 2010.

[27] S. Zhang, Y. Zhu, Y. Wang, W. Zhang, and Y. Li, "A Demographic-based and Expertise-enhanced Collaborative Filtering Method for E-government Service Recommendation ⋆," vol. 6, pp. 2463–2480, 2014.

[28] Said, A. et al., A Comparison of How Demographic Data Affects Recommendation. , pp.1–3.