Enhancing Recommender System for Matrimonial Sites using Collaborative Filtering Method

Sampath.M.K, Nithya.C, Mohana Priya.R
Assistant Professor
Department of Computer Science and Engineering,
Narasu's Sarathy Institute of Technology,Salem, India
mksam83@gmail.com, nithya.acet@gmail.com, mohanapriyar25@gmail.com

Abstract- Recommender Systems helps users to find items of interest from a large number of available items. Collaborative Filtering is the commonly used technology for recommender systems. The role of recommender system in matrimonial sites is profile matching based on the preferences given by the users. The users of matrimonial sites have a problem of overloaded choices of partners. This is because the currently used collaborative-filtering-based recommender systems focus only on the information about the way the users interact with the system and the original interests of the user is not identified. In this paper we provide a collaborative-filtering-based recommender system that identifies the user's latent interest and provides top-n recommendations to the active user based on preferences and interests of the user. The top-n recommendations are identified by ranking recommended profiles in terms of weightage. This helps the users of matrimonial sites to easily identify a perfect match. Moreover opinion mining can be used to identify profiles which are of interest to maximum users. This explores the minds of current generation as to what kind of profiles the users are mostly interested in.

Keywords-Recommender System, Collaborative Filtering, Opinion Mining.

I. INTRODUCTION

One of the key problems in today's world is to deal with lots of information. The information is available to the user whether or not he looks for it. This growth of information on the internet led to the development of recommender systems. The Recommender System (RS) attempts to reduce this information overload and helps individuals to make decisions. The main goal is to recommend items based on the preferences of the end user. Here 'item' is a general term that refers to what the system recommends to the user. The main task of the RS is to predict items of user interest and then recommend them. The popular existing recommender systems are Amazon.com for e-shopping, MovieLens for movies etc. Many different types of RS have been developed in recent years. Most of the existing RS uses the Collaborative Filtering approach.

A. Collaborative Filtering

The collaborative filtering [1] makes automatic predictions (filtering) about the interest of individual by collecting preference information from many users (collaborating). There are model-based collaborative filtering and memory-based collaborative filtering. The Model-based CF makes predictions on real data by learning a model. It then makes recommendations by computing the expected value of user prediction based on his ratings on other items. Memory-based CF uses user rating data to compute similarity between users and then recommendations are given. This system first identifies a set of users called neighbours. Then it uses different algorithms to combine preferences of neighbours to provide recommendation to the active user. It is also known as user-based collaborative filtering and is most widely used in practice.

B. Opinion Mining

Opinion mining also known as sentiment analysis is used to easily identify the attitude of speaker or writer. The source of data is the opinions given by various people on a particular thing. The opinions can be positive negative or neutral. The basic components of opinion are opinion holder, object and opinion. Opinion holder is the person who holds the opinion; object is the one on which opinion is given and opinion is the attitude of the opinion holder towards that object. Based on the opinions of people on a particular object, we can identify the user's interest on that particular object.

The users of matrimonial sites are facing a problem of large matches about the partner. This is because, the recommender system used currently will provide recommendations based on the users preference given in their profile. Hence there will be large number of profiles matching a user. In this paper we propose an enhanced collaborative filtering for RS that make recommendation based on both user's preference and user's interest. As a result the user will have a limited number of choices from which he can select a match easily.

II. RELATED WORK

A lot of work has been done in the field of recommender systems. It is a popular way of solving the information abundance problem. The recommender systems that exist currently are Amazon.com, Tripadvisor, Netflix etc. In matrimonial matchmaking, RS is an emerging research area with lots of scope. As the users find easy to search partners in their busy lifestyle, matrimonial matchmaking is becoming more popular.

A. Matrimony Matchmaking Recommender System

As matrimonial matchmaking is an emerging area, only few papers are available for online dating. Recon [2] is a system that uses content based approach by implementing compatibility and reciprocal algorithms. The system used both user profile and interactions. It extracted the user's implicit preferences by inferring the user's interaction with other users and then matches them with profiles.

III. OVERVIEW OF MATRIMONIAL SITE

First, the users of matrimonial sites have to create their own profiles by registering into it and giving details willingly. They must also provide preference details about the partner. User profile will be generated once the user submits the registration form. The active user's preferences about their partners are taken and matches are identified using item-based collaborative filtering. Then user interests are identified by using personalization technique and profiles that match user interests are identified. Finally the profiles that satisfy both user preference and interest and are considered for recommendation and then top-n profiles are recommended based on weightage of profile to each active user. One of the main issues of collaborative filtering is overspecialization problem. The issues have been overcome in this paper.

User's profile is separated into two groups of male and female profile. The union of male and female profile will be the total number of users of the matrimonial site. The profile of each user is further divided into account information, contact information, personal profile and preferred partner's profile. Account information comprises of login name, password, users name, date of birth and security question and answer. Contact information contains address, city, state, country, mobile number and email-id. Personal profile consists of attributes like age, height, religion, caste, education, occupation, income, complexion, body type, diet, smoke and drink. The preferred partner profile consist of user's expected range of age, range of height, range of income, religion, caste, education, occupation, complexion, body type, diet, smoke and drink.

The system allows user to specify 'doesn't matter' for any of the attribute. For example, if the user specifies 'Doesn't matter' for education attribute, then the user allows partner for all education levels. The system also allows user to select more than one value for attributes like Education, Occupation, Complexion, etc. There are some preference attributes that may vary for male and female users like Smoke and Drink which is preference attribute only for female users.

IV. PROPOSED RECOMMENDER SYSTEM

A. Recommender System

The main function of RS is to increase user satisfaction, increase user fidelity and better understand what the user wants. This can be achieved by collaborative filtering approach which is most widely used in existence. The technique used for CF is item based collaborative filtering and user interest expansion. The recommender system takes as input the preference attributes value of each active user and recommends profiles that matches the user's expectation. All the attributes are not equally important. Based on the importance the attributes are divided into two groups. One group consist of attributes of high importance like religion, caste, occupation, physical status, marital status, diet, smoke and drink. Other group consist of attributes of less importance like age, height, education, income, body type and complexion. The RS will provide a list of recommendation where all the attributes of high importance matches perfectly but the attributes of low importance may or may not match perfectly.TABLE I shows the preference attributes and some possible values for those attributes.

TABLE I:Attribute, Importance and Possible Values

Sl. No	Attribute	Importance	Possible Values
1	Age	Low	21 to 35
2	Height	Low	4ft 5 inches to 6ft 5 inches
3	Marital status	High	Unmarried, Widow/Widower, Divorced
4	Physical Status	High	Normal, Physically challenged, Doesn't matter
5	Religion	High	Hindu, Christian, Muslim, Jain, Sikh, Buddhist, Parsi
6	Caste	High	Brahmin, Vysya, Mudaliar, Gounder, Chettiyar, Catholic, Protestant, Shia, Sunni, Digambar, Shwetambar
7	Education	Low	PhD, Masters in Engineering, Bachelors in Engineering, Masters in Arts and Science, Bachelors in Arts and Science, Masters in Medicine, Bachelors in Medicine, Masters in Legal, Bachelors in Legal, Financial Qualification like CA, ICWA, Service like IAS, IPS, Diploma, HSC, Below HSC
8	Occupation	High	Doctor, Nurse, Software Professional, Hardware Professional, Lawyer, Professor, Lecturer, Teacher, Researcher, Architect, Business, Social worker, Journalist, Media Professional
9	Annual Income	Low	1,00,000 Rs. to 15,00,000 Rs.
10	Body type	Low	Slim, Athletic, Average, Heavy
11	Complexion	Low	Very Fair, Fair, Wheatish, Dark
12	Diet	High	Veg, Non-Veg
13	Smoke	High	Yes, No
14	Drink	High	Yes, No

B. Item Based Collaborative Filtering

Item based collaborative filtering is a technique which identifies neighbours of a user and makes similar recommendations. The neighbours are identified as users who have similar preferences. For instance, a new male user has registered into the matrimonial site and he has specified partner preference attributes value as follows:

Age: 23 to 26, Height: 5ft 3 inches to 5ft 6 inches, Marital Status: Unmarried, Physical Status: Normal, Religion: Hindu, Caste: Brahmin, Education: Bachelors in Engineering, Masters in Engineering, Masters in Arts and Science, Occupation: Software Professional, Annual Income: 4, 00,000 Rs. to 5,00,000 Rs., Body Type: Average, Complexion: Wheatish, Fair, Diet: Veg.

Based on the importance of attributes the neighbours of this user are the male users who have same preference for Marital Status, Physical Status, Religion, Caste, Occupation, and Diet. Now the profiles which were recommended for those male users will be considered for recommendation to the new user. Then the less important attributes are compared with the profiles considered for recommendation. The profiles which have maximum similarity are considered first than profiles with less similarity. This is how the Item-based collaborative filtering works.

C. Identifying User Interests

The necessity of identifying user interest is to better understand the user's requirements because user interest may better connect user and partner. Moreover the interests can change from time to time. If the recommender system cannot identify these changes and make only recommendation based only on new preferences then it leads to overspecialization problem. It can be overcome by identifying the time-to-time change of interest. User interest can be identified by keeping track of user's previously visited items. These items can be used to identify interest of a user and recommend related profiles. Here the item represents profiles which the user has visited and the latent interest here is the value of the attribute which the user is interested. In the profiles the latent interest can be viewed as any one of the attribute value. So the RS will recommend profiles that have this specific attribute value. Since each profile has many attributes, it belongs to many latent interests. At the same time, different profiles may have similar attribute values; they may refer to one same latent interest.

D. Top-N Recommendations

If more number of individuals uses a matrimonial site, then it is obvious that there are chances for the recommender system to recommend more profiles to each user. To avoid this scenario we go for the top-n recommendation. In order to identify the top-n profiles for recommendation we must rank the profiles considered for recommendation based on some measure. Here weights are assigned to each based on his/her interest and preference. Fig.1.Represents the overall architecture of the proposed Recommender System.
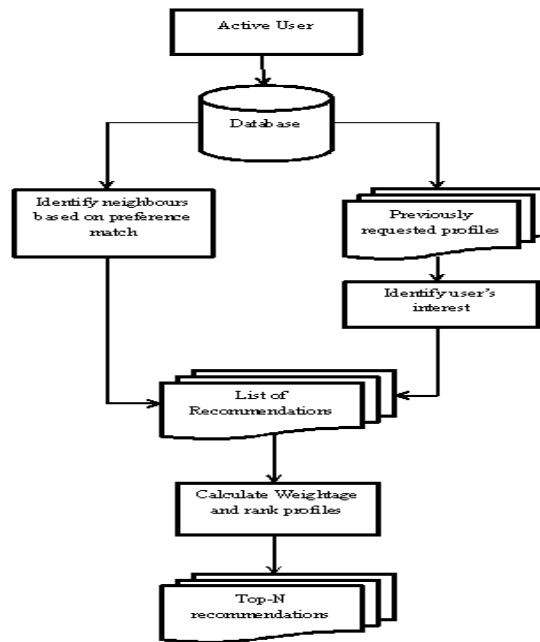

Fig. 1. Architecture of proposed Recommender System

V. TREND EXPLORATION BY OPINION MINING

Opinion mining is used to identify the attitude of user towards different profiles. Although the recommender system recommends only top-n profiles, all the active users will not be interested in all the profiles. Since each male/female profile is recommended to more than one female/male users, each profile is viewed by many users. But the users will not give request to all the recommended profiles. The user will give request only if he/ she have a positive opinion on the profile. When each active user gives request to each profile, the request count of the profile is incremented. The male profiles which have maximum request count are identified and the female profiles with maximum count are identified.
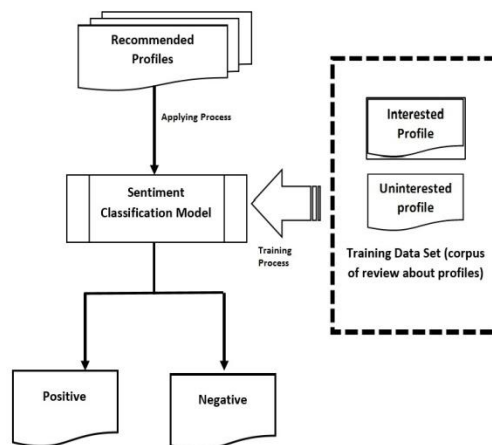

Fig. 2. Trend Exploration

Fig. 3. Represents opinion for system recommended profiles. Each active user's preference is checked with these identified profiles. If the important attributes like age, religion and caste are matched, then this profile is also recommended additionally to the active user. Fig. 2. represents the trend exploration. This helps us to explore the common interest of users as to what kind of profile they expect. It also reveals the minds of the people of current generation.
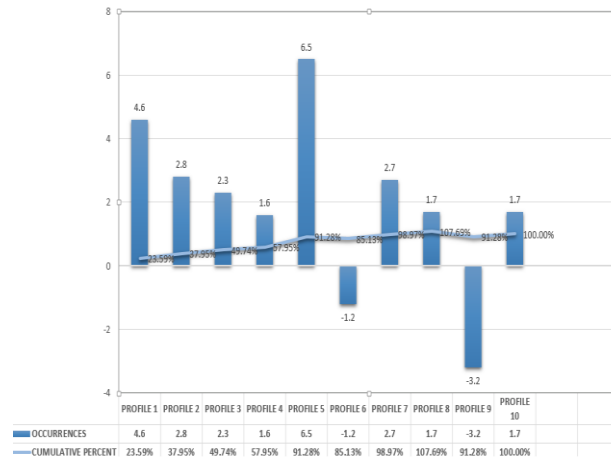

Fig. 3. User Opinion for System Recommended profiles

VI. CONCLUSION

In this paper we have proposed a recommender system that uses collaborative filtering approach for making recommendations. The recommender system will identify profiles to be recommended based on both item-based collaborative filtering and by identifying user interests. It makes top-n recommendation by assigning weightage to each attribute and finding the total weight of each profile. By ranking the profiles, top-n profiles are recommended to the active user thus helping them to easily find the desired partner.

REFERENCES

[1] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proc. 10th Int. Conf. WWW*, 2001, pp. 285–295.

[2] L. A. S. Pizzato, T. Rej, T. Chung, I. Koprinska, and J. Kay, "Recon: a reciprocal recommender for online dating," in RecSys, pp. 207-214, 2010

[3] Smita Khavate, Mrs. M. A. Potey"Partial fuzzy recommender system for Indian matrimony"

[4] Qi Liu, Enhong Chen, Chris H. Q. Ding Enhancing "Collaborative Filtering by User Interest Expansion via Personalized Ranking"

[5] Francesco Ricci · Lior Rokach · Bracha Shapira ·Paul B. Kantor "Recommender Systems Handbook" pg 107 – 140

[6] Michal Holub, Maria Bielikova "Estimation of User Interest in Visited Web Page"

[7] Feng Qiu, Junghoo Cho "Automatic Identification of User Interest For Personalized Search"

[8] Tamil Matrimony .com

[9] S. Debnath, N. Ganguly, and P.Mitra, "Feature weighting in content based recommendation system using social network analysis," in *Proc. 17th Int. Conf. WWW*, 2008, pp. 1041–1042.

[10] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," IEEE Trans. Knowl. Data Eng., vol. 17, no. 6, pp. 734-749, 2005.

[11] L. Brozovsky and V. Petricek, "Recommender system for online dating service," in Proceedings of Znalosti 2007 Conference, Ostrava), VSB, 2007.

[12] R. Nayak, M. Zhang, and L. Chen, "A social matching system for an online dating network: A preliminary study," in Proceedings of the 2010
IEEE International Conference on Data Mining Workshops, ICDMW '10, (Washington, DC, USA), pp. 352-357, IEEE Computer Society, 2010.

[13] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Foundationsand Trends in Information Retrieval*, vol. 2, nos. 1–2, 2008, pp. 1–135.