FRIEND RECOMMENDATION SYSTEM FOR SOCIAL NETWORKING SITES USING CLOUD COMPUTING ENVIRONMENT

Sumit Kumar Sharma

sumit.sharma1902@gmail.com
Research Scholar
Sri Aurobindo Institute of Technology, Indore

Manisha Lokhande

manisha.lokhande@sait.ac.in
Assistant Professor
Sri Aurobindo Institute of Technology, Indore

ABSTRACT

With the changing in internet, way of using it also changing. Internet not only just provide a way of interaction through operating system but also involving in different fields like artificial intelligence, machine learning etc. Social networking site is mostly used and popular platform of internet. It creates connectivity among people with similar features. Internet service in social networking is making it essential in life of people. Keeping in mind about user nature and interest makes it easy to recommend similar characteristic friend. It also provides a way of enhancing business, promoting products, getting current new, updates etc. It helps to be in touch with our contacts by recommending them similar characteristics of user.

This approach is very innovative to find friends of there interest. In our work we have used clustering, filtering approach. Then, DBSCAN is used which forms cluster of similar user based on there features. To find more accurate result collaborative filtering based on indexing is used. Recommender system search most relevant friend among many. With the change in demand of user, system also changes. Data mining is also used in our work.

Keywords: Friend Recommendation, Data Mining, Cloud computing

1. INTRODUCTION

With the rising of internet and fast life everyone want on demand fast accessing of objects. More data is uploaded with the increase of using internet. With one search user gets millions of search result in which he gets confused to refer the one. Many queries also arises in which user did not get relevant search result. This unwanted situations results in poor performance of system and data. Then recommendation system is used to solve the issues. It is a tool which is used to obtain relevant information. This relevant information can be examined from the similar characteristics of user.

Machine learning has the ability to take decision, it is a kind of artificial intelligence. It has a capability to develop computer programs and make changes according to the change in data. Whereas, Data mining technique extracts knowledgeable information from the collection of data. They both can be differentiated on the basis of analysis approach and prediction approach.

Our work is based on the tool which is used to find the interest of user and then convert it into

characteristic form. Then data mining technique is used in which data is converted into knowledgeable form using the mining tool. Recommendation system use the result of mining into specific way, it is used in many applications like social networking etc.

1.1 Data Mining works under the following parameters :

1. Data warehouse : Collection of heterogeneous data from all possible source is called data warehouse.

2. Data cleaning : Removal of unwanted inconsistency from the collection of source data is called data cleaning.

3. Data preprocessing : Process of extracting knowledgeable information from the collected data is called data preprocessing.

4. Data transformation : Process of converting data into homogeneous form is called data transformation.

5. Knowledge extraction : Searching of hidden information using processed data and extracting knowledge from that data is called knowledge extraction.

1.2 Recommendation System works under the following techniques :

1. Knowledge based recommendation system : Decision rules and patterns are taken into consideration for providing recommendation. It is mainly responsible for recommending user. The recommendation system which mainly works on changes is better because it will recommend better results.

2. Content based recommendation system : This system uses the history and extract the information from the collected data. The content works on the popular data which is mostly used. Sorting is performed for the ranking of data and that ranked data is used for recommendation, which is based on content.

3. Collaborative based recommendation system : This approach is similar to content based with little difference. The content of the collaborative approach is refined using similar group of user. Keeping in mind of the similarity among user, it is classified.

4. Demographic based recommendation system : It is used for classifying user on some certain characteristic. Characteristic can be of any type depending on there choice, features of similarity etc.

2. RELATED WORK

Wang et al. In[1] proposed the recommendation system based on machine learning and data mining which can perform the convenient role of friend recommendation. User features and activities are extracted and analysis of it as an input source is done for semantic approach for recommendation. It mainly focuses on current activities not the complete behavior of user. In social networking user connects with friends on the basis of similarity in characteristic and interest. This attempt is based on the closeness of user with another user.

Kwon et al. In[2] describes about the recommendation system which is always preferred by the user of social networking site because of large volume of data of friends. Algorithms like context aware are used which helps extracting physical and social nature of user. Hybrid approach is used depends on features like physical and context. Recommendation algorithm is more preferred because it give more accurate results.

Bian et al. In[3] by using collaborative filtering developed a solution for user based on there personality which concluded that on the basis

of personality user interest rate can be obtain. Which helps defining nature of friend and group of friend. If user is sharing same place, same area, common background, similar friends, then recommendation will become more easier and. This creates a matching graph based on personality and other common elements. Which will be a good way of finding relevance information. So as to get information about data mining approach and recommendation system, relevant algorithm is performed

Linden et al. in[4] studied about recommendation system for large dataset and solution which are proposed. That recommend products which are based on characteristics of items. They concluded about nature of every user which is unique. Previous transactions also defines nature of products while shopping. Integrated way of recommendation is very effective in shopping purpose. Both the clustering approach and filtering approach are used for recommendation purpose.

Data mining study define about clustering approach, which is a good approach for social networking so that they can prepare cluster of similar users.

Kanungo et al. in[5] defines about the implementation and understanding of K-mean algorithm which strongly justify the use of this approach for social networking sites. The only issue arises with this proposal is recommendation system, it can not be used as recommendation system and manage to prepare cluster of similar elements. A matching graph is considered which provides solution for the purpose of friend recommendation.

Kacchi et al. In[6] proposed a friend recommendation approach which is based on graph called matching graph. He concluded about innovative lifestyle approach on which

similarity can be checked and from groups belonging to that similar feature. Friend recommendation is developed to justify the solution based on matching graph. This graph technique is based on mesh mapping, which finds similar score with respect to existing user. Heavy computation is done in this work of recommendation . Mesh matching is not a sensible solution for large data. The complete study observes that friend recommendation is a requirement of technology.

1. DATA COLLECTION AND PREPARATION

Data sources and information is very important to justify accuracy of system. Data is collected in two ways one from direct user or other from existing system. Collection of primary data can be done by survey which is a good way. Collection of secondary data can be achieved from pre-existing dataset. Here, in this work primary data collection is collected as the prime source. General schema is defined for user, and different types of user are involved in it, different profession, cast etc. Predefined schema is used for the collection of data consisting 43 attributes from many different category. After it data cleaning procedure is applied, which removes the tuples which are incomplete from data source. Afterwards, stop words removal and lemmatization module is developed which helps to improve the quality of data source. Then, Tokenization scheme and abbreviation removal is applied to break complete work into single values and extract to get more accurate lifestyle answers.

2. CLUSTERING APPROACH

Clustering is an approach to classify all elements in such a way that every similar element should be resided into single group

based on their similarity. Subsequently, it also reside irrelevant elements into another group based on their similarity value and maximum cluster size. Here, K-Mean clustering approach has been used to construct group of similar users based on lifestyle similarity. It is one of the simplest unsupervised learning algorithms which simplify the work of mining by classifying the similar elements in cluster using k-centroids parameter. It calculates distance between each element to evaluate similarity and reside them into single cluster by comparing with k-centroid parameter.

The challenge faced in this section is to map the similarity of each user and transform them into quantify figures. Quantification technique is implemented to convert all relevant values into matching score.

After it, the output come is forwarded to the module of clustering for cluster making. which generates the most similar users based on relevance distance.

3. RECOMMENDATION APPROACH

Recommendation system uses a filtering method for rating, similarity or preference score are predicted to get the frequency between item and elements. Recommendation systems has increased popularity in recent years due to its use in wide area of applications like movies, books, articles etc. A customized recommendation algorithm developed and used with K-mean clustering algorithm so that more accurate and relevant solution can be achieved. For input data source recommended cluster is used and similarity score is calculated.

Similarity score indicates total lifestyle of each user. In simple form, it can be concluded that high similarity score indicates

more lifestyle closeness when comparing with users having low similarity score.

Similarity threshold value is used to filter the retrieved id and recommend most close references. High threshold indicates high filter strength and will recommend users with close values. Low threshold indicates low filter strength and high numbers of users.
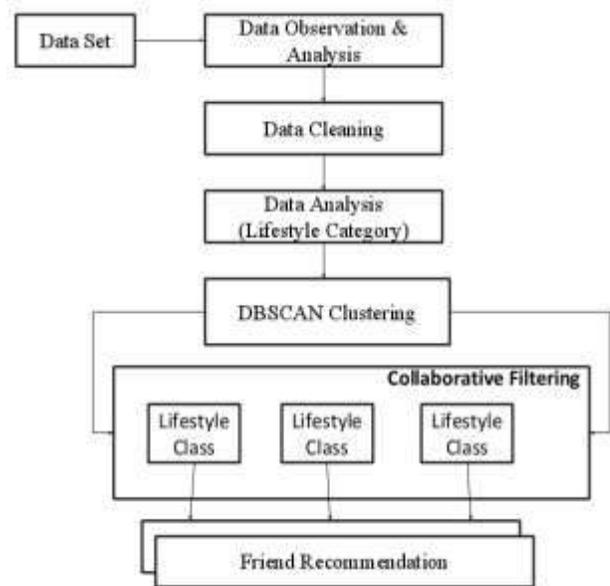

Figure 1 : Proposed Architecture

4. EXPERIMENT ANALYSIS

Recommendation tool developed is based on java. Implementation proposed is classified into four modules. All the collected data is exported into .CSV file format from Google docs. Then is loaded to perform parsing process. Removal technique is applied on incomplete data and is implemented for cleaning of data purpose. Afterwards, stop word removal, lemmatization and tokenization purpose is achieved. This module helps to exact and get more accurate data source for clustering process.

The challenge faced during clustering process was mapping of all user life style into numeric form. A quantification process is used

to transform all sentiments into numeric figures. K-mean clustering approach is performed to extract similar users based on similar information. Clustering help to retrieve relevant users.

A recommendation technique based on collaborative filtering has been implemented. Similarity weight is considered and evaluated then ranking of user in similarity index clustered is formed. It is the total sum of all weights estimated during quantification process. Lastly, threshold value is used to filter retrieved document and generate most similar users as final recommendation.

5. RESULT ANALYSIS

The complete phenomena will be evaluated on basis of Recall [Accuracy], Precision and F-Score [Final Score]. Initially ten users have been selected on random basis for friend recommendation purpose. Afterwards, five different threshold values have been selected to obtain results.

In this work, Recall and Precision parameters will be calculated for each selected users based on variable threshold values. A comparative study of all recall and precision will be done. Afterwards, f-score will be calculated to measure the overall performance of complete system. Average recall and Average precision of each threshold scenario has been considered to derive single f-score.

6. CONCLUSION

The complete work concludes friend recommendation system which recommends friends on the basis of there behavior, nature and characteristics. A modified clustering and filtering approach is proposed as the hybrid solution for recommendation purpose. A

customized quantification and filtering approach is performed to simplify and get accurate performance. The performance is calculated on the basis of recall, precision and f-score.

REFERENCES

[1] Zhibo Wang, Hairong Qi, "Friendbook: A Semantic-Based Friend Recommendation System for Social Networks," IEEE Transactions on Mobile Computing, Vol. 14, No. 3, MARCH 2015.

[2] J. Kwon and S. Kim, "Friend recommendation method using physical and social context," Int. J. Comput. Sci. Netw. Security, vol. 10, no. 11, pp. 116–120, 2010.

[3] L. Bian and H. Holtzman, "Online friend recommendation through personality matching and collaborative filtering," in Proc. 5th Int. Conf. Mobile Ubiquitous Comput., Syst., Services Technol., 2011, pp. 230–235.

[4] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: item-to-item collaborative filtering," IEEE Internet Computing, vol. 7, Jan. 2003, pp. 76-80.

[5] Tapas Kanungo, Nathan S. Netanyahu, "An Efficient k-Means Clustering Algorithm: Analysis and Implementation". IEEE transactions on pattern analysis and machine intelligence, ol. 24, no. 7, july 2002.

[6] Tofik R. Kacchi, Anil V. Deorankar, "Friend Recommendation System based on Lifestyles of Users". International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB16). 2016 IEEE

[7] B. Liu, W. Hsu, and Y. Ma, "Integrating Classification and Association Rule Mining," 20th Int. Conf. on Knowledge Discovery and Data Mining (KDD), New York, USA: 1998.

[8] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," ACM SIGMOD Int. Conf. on Management of Data, New York, USA: 2000.

[9] S. Morinaga, K. Yamanishi, K. Tateishi, and T. Fukushima,"Mining product reputations on the Web," 8th ACM Int. Conf. On Knowledge Discovery and Data Mining (KDD), Edmonton, Canada: 2002.

[10] C. Scaffidi, K. Bierhoff, E. Chang, M. Felker, H. Ng, and C. Jin,"Red Opal: product-feature scoring from reviews," 8th ACM Conf. on Electronic Commerce (EC), San Diego, California: 2007.

[11] K. Kageura and B. Umino, "Methods of automatic term recognition: A review," Terminology, vol. 3, Jan. 1996, pp. 259-289.