International Journal of
Emerging Research in Management &Technology
ISSN: 2278-9359 (Volume-5, Issue-3)

Research Article | **March 2016**

Social Recommendation System Technique: Survey

Aniket Pote[1], Swapnil Sutar[2], Priyanka Sonawane[3], Fatima Inamdar[4]
[1, 2, 3] Department of Computer Engineering, VIIT, Maharashtra, India
[4] Assistant Professor, Department of Computer Engineering, VIIT, Maharashtra, India

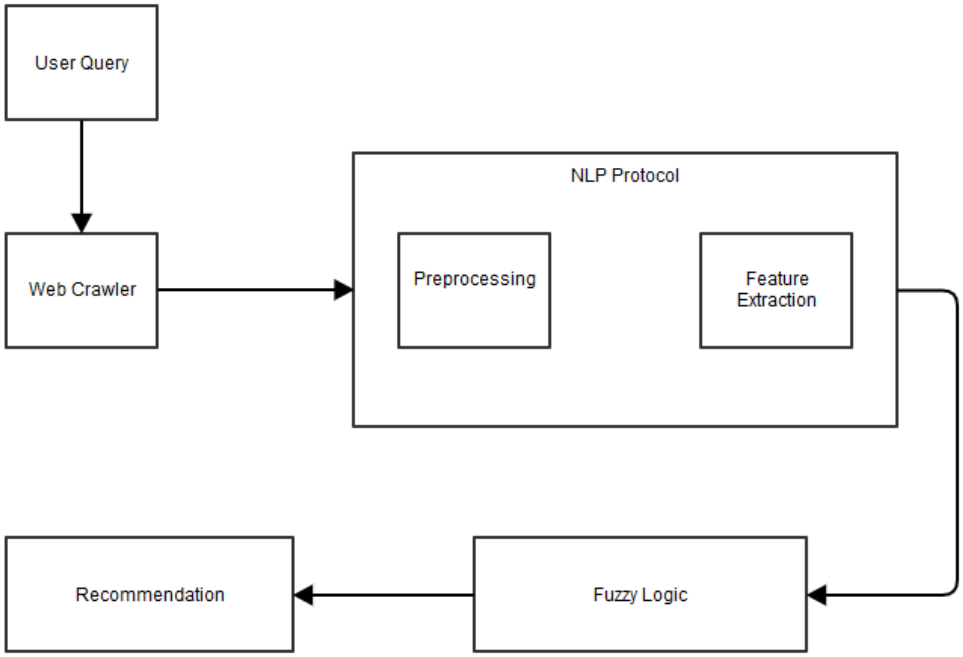
Abstract—

In the current scenario, social networking sites are becoming an inseparable part of the user's life. Every day life, the users or people on social networking sites are much interested to find the similar users for their current requirements. So many of the systems recommendation individuals based on their profile taste which is created during registration, but this way it fails to consider individual's current requirements. There are techniques used in the recommendation systems like collaborative filtering for their implementation. So the problem with these systems is that they fail to consider user's current interests. As a solution to this, we proposed a new recommendation system, which considers user's current interests and provides results that are more accurate.

Keywords— Recommendation system, web crawling, preprocessing, feature extraction, Fuzzy logic

I. INTRODUCTION

We are proposing personalized recommendation using social circle and user interest.



First step of this is use of web crawler. It is process of search engine combing through web pages in order to properly index them. In NLP (Natural Language Processing), it includes preprocessing and feature extraction. In preprocessing removal of stop words, special symbol removal, stemming is done. In feature extraction, it includes numeric and top word extraction for our proposed system. In numeric data extracted from site, user fired query is extracted to be used as crisp value to be given to fuzzy logic for evaluation of result which is recommendation. The next phase is application of fuzzy logic. In fuzzy logic, the inputs are provided by the NLP (Natural Language Processing) protocol phase, which are crisps values. Crisp values are used in the process of fuzzy logic or in fuzzy algorithm, for providing results, which is actual recommendation. Next phase provides recommendation. By using the fuzzy logic, it can give the recommendation. According to priorities assigned to the comments in relevance with users fired query, recommendation is generated by the system in decreasing order of relevance of the comments to the user fired query and display to the user.

II. LITERATURE SURVEY

[1] In the current scenario more and more users want to share their experiences and their views. The Friend circle based new social factors like Interpersonal Interest and Influence bring challenges to solve sparsity problem and cold start problem in data sets. This paper uses three factors like Personal interest, Interpersonal Interest and Influence to provide recommendations using matrix factorization. The drawback with this approach is, it does not give real time results as it considers the probability of occurrences.

[2] To find interesting and relevant item or also the product, a tool is used which is Recommendation System. In today's scenario, people want to share their reviews, ratings or in all experience which is then used to recommend user the items of his or her interest. Potential growth of social networking sites leads to huge amount of data. Personalized Recommendation Systems are used to handle this overwhelmingly large amount of data. This paper considers the influence factors for the Personalized Recommendation System, which are: importance of Recommendation System, different methodologies and social factors. This paper describes different methodologies used to build recommendation system, which are considered along with user's personal and interpersonal interest by base matrix and social matrix factorization techniques on social network.

[3] High volumes of data can be processed with the help of data mining process; whose applications has dataset with high dimensionality. It degrades the performance of the machine learning. The problem of high dimensionality is solved using technique called Dimensionality Reduction (DR), which reduces high dimensionality. Feature extraction is important technique in Dimensionality Reduction, which is used for extraction important features. This paper considers various feature extraction methods and the issues related to these methods. Hence, future scope is to find issues related to it these algorithms and propose new algorithm, which improves classification accuracy as well as extracts new features.

[4] Automated process of extraction of useful data from large datasets or databases is Data mining. Real world applications contain lot of erroneous data. To rectify the erroneous data present in the application, an important technique present in the data mining is data preprocessing. A lot of data mining applications contain data, which is in high dimension, and this high dimensionality reduces the mining algorithm performance. It also increases the space and time required for processing the data. The issue of high dimensionality is resolved using the Dimensionality Reduction (DR) technique, which is divided into feature selection and feature extraction. This paper carries out a detailed survey to know how 2 techniques are used to solve problem of Dimensionality Reduction. Also to extract most relevant features various statistical measures are explained. To extract new set of features from the original features, different statistical techniques are analyzed.

[5] Information Retrieval uses various tasks to increase the recall rate and giving most relevant results. Stemming is one of the processes used in the Information Retrieval System to reduce word to its root form or stem. From the manual to automatic methods, the stemming can be performed by a number of ways with each having its own advantages. These methods can be from language specific to language independent. This paper represents a comparative study of the available alternatives for stemming which are widely used in Information Retrieval System for increasing its effectiveness and efficiency. Still there are some issues that need to be dealt properly. Sample size considered in statistical stemming is under debate. From smaller sample size, stemming will be faster but the language coverage will be in doubt. Nevertheless, in the case of the large samples stemming will take long time.

[6] Due to large number of increasing E-commerce sites there is been an information overload on internet which causes difficulties for the customers in finding and purchasing the right products best suitable for their needs. Fuzzy Logic technique is used in the personalized recommendation system, which is proposed in this paper. This system mines information intelligently about the features of the laptop computers and provides the services to potential buyers by considering their personal needs and preferences for recommendation. To measure the similarity between the product features and consumer needs a concept called Fuzzy Near Compactness (FNC) is employed which helps in the recommending the optimal products to potential buyers. Proposed system's experimental results considered 50 laptops from provides Acer, Dell, HP, Sony and Toshiba. The proposed recommendation system is a solution for less frequently purchased products. Different techniques are used in the proposed system for mining customer requirements and product details which provides the optimal recommendation to the potential buyers. The system provides information about product to the buyer that best satisfy their needs. The system also has the potential of increasing the online

[7] Many E-commerce applications today use Recommender Systems that provide user with the recommendation that the users may prefer or they can also provide predictions for how much a user would prefer each item. Collaborative Filtering and Content Based Filtering are the 2 common approaches for providing recommendations. By combinations of these two approaches, a hybrid system can be developed which considers rating of user and item's features. This is used to recommend items to the user. Existing data analysis tools can be used for limited amount of data but for the large databases Hadoop is used which is a software framework for distributed processing of large data sets. The proposed system is more fault tolerant than the existing as it uses rating from the user to predict interest and also the analysis of the item is done to find the features. The proposed system is adaptive which frequently updates rating list and also finds the users updated interest. According to a survey, more than 60% of the users purchase the recommended products. In this way, more success can be achieved in providing relevant recommendation by incorporating customer interest. By analyzing feature of the product and matching it with the user interest, recommendations can be provided more accurately. The proposed system can be used for recommending books to the user but, can be expanded for movies recommendation. For dealing with website recommendation, total number of view for website should be considered as a metric for providing efficient and accurate recommendation.

[8] Recommender system are useful tools, which provide web users a adaptive web environment. In these days of the E-commerce technology, a big challenge is having a user-friendly website. In this paper, taking into consideration both the collaborative and content based filtering features, a technique is proposed by presenting a fuzzy recommender system. This system is based on the collaborative behavior of ants (FARS), which works into 2 phases namely modeling and recommendation. In modeling, the user's behavior is modelled offline and in the recommendation, the results from the first phase are used for online recommendation. The ant-based algorithm provides the optimal solution whereas the possibility of capturing the uncertainty among user interest is provided by the fuzzy techniques. The performance

evaluation of the FARS is achieved by the use of log files of "Information and Communication Technology Centre" of Isfahan municipality in the Iran and which is then compared with ant based recommender system (ARS). The results of this system stated and proved that the fuzzy Ant approach provides more functional and robust approach.

[9] Typically the data has to be in standard from the techniques for learning from the data. The measurements must be in specific format and for classification, a specific goal should be specified. Some databases are in standard form while some others are not. The significant efforts are dedicated for converting data into suitable features. This paper describes the procedure of automated feature extraction, which is adapted from the modern text categorization technique. This maps the large databases into manageable datasets converted in standard form. An application is considered to mine large database of home appliances service records for major retailers.

[10] A common preprocessing step in Text mining application is stemming. It is also a very common requirement of the NLP (Natural Language Processing) functions and it is very important in Information Retrieval System. Stemming's main purpose is reducing different grammatical words into their root forms. We have considered different method of stemming and their comparison. This comparison is related to the usage, advantage, limitations and also the difference between lemmatization and stemming is discussed. This paper discusses different algorithms which gives the conclusion that not one algorithm gives 100% output. That is if one algorithm does better work at one area, the other does better work in another area. For efficient stemming, there is need for method and system. It should reduce the heavy tradeoff between false positive and false negative. To reduce the stemming errors stemmer using syntactical and semantical knowledge should be developed. A good lemmatize development could help in achieving the goal.

[11] In this paper, Feature extraction for the classification in the data mining process, the data mining process has important step of dimension reduction. It takes into account the problem of "curse of dimensionality" and considers feature extraction as classification task to overcome this problem. Three feature extraction methods and three applications with respect to classification tasks are considered. The results give the accuracy schemes with conclusion about most appropriate feature extraction method. The knowledge needed to integrate the processes of feature extraction and classification is considered and decision support system to integrate these processes is taken into account with goals and requirements of the decision support system. To build up the decision support system the means of knowledge acquisition are considered.

[12] Data mining process has a number of important steps one of which is data preprocessing which is an important and critical step. It has huge impact on the data mining project. Data preprocessing is a process in KDD (Knowledge Discovery in databases) which offers better conditions in subsequent analysis and reduces the complexity of data. It helps in understanding the nature of data and it also helps in performing data analysis accurately and efficiently. There are number of tools and methods used for the preprocessing like sampling, transformation and removing noise. This method is used for association rules algorithms like Apriori Algorithm.

III. PROPOSED METHODOLOGY

The proposed systems idea is realized from the aspect that online social networking site users are often getting very bad recommendations for their searched queries. This is due to the old techniques or also the old posted comments of the users on the web pages. In order to overcome this, we are proposing a technique for providing a fresh recommendation to the users. These recommendations are for their searched keyword on the online social networking sites. For its realization, we are creating an enriched online social networking site that will run in the LAN, where users allowed to post the comments. Whereas on the other hand, our system will recommend the right users for their fresh posts on the web pages. In paper, we discussed various techniques for providing the personalized recommendation.

The different modules, which can be used in the system, are:

1) Crawling: The Crawling is used to scan the data present on the web page which is to be used.
2) NLP Protocol: NLP protocol is used for many purposes, which are preprocessing and feature extraction.
3) Fuzzy Logic: This is to provide final recommendation using the crisp values taken as from the NLP protocol module.

IV. CONCLUSION

We have studied various techniques to propose a new technique, which provides better social recommendation than the existing recommendation systems context with the relevance with user, fired query. This technique can be used by the recommendation systems to improve its accuracy and provide most relevant results for the user's query.

REFERENCES

[1] Xueming Qian, He Feng, Guoshuai Zhao, Tao Mei," Personalized Recommendation Combining User Interest and Social Circle", *IEEE TRANSACTIONS KNOWLEDGE AND DATA ENGINNERING* VOL:26 NO:7 YEAR 2014

[2] Anil Rathod, Indirmma M," *S Survey of Personalized Recommendation System with User Interest in Social Network*" Anil Rathod et al, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (1), 2015, 413-415

[3] N. Elavarasan, Dr. K. Mani, "*A Survey on Feature Extraction Techniques*", International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297:2007 certified Organization) Vol 3, Issue 1, January 2015

[4] V.Arul Kumar, N. Elavarasan, "A *Survey on Dimensionality Reduction Technique*", International journal of Emerging Trends & Technology in Computer Science(IJETTCS), Volume 3, issue 6, November-December 2014 ISSN 2278-6856

[5] Deepika Sharma, "Stemming Algorithms: *A Comparative Study and their Analysis* ", International Journal of Applied Information Systems (IJAIS) – ISSN: 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 4– No.3, September 2012

[6] Ojokoh, B. A., Omisore, M. O, Samuel, O. W, and Ogunniyi, T. O.,"*A Fuzzy Logic Based Personalized Recommender System*", IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN: 2249-9555 Vol. 2, No.5, October 2012

[7] S. Vinodhini, B. Govindarajalu, V. Rajalakshmi, "*Building Personalised Recommendation System With Big Data and Hadoop MapReduce*", International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 3 Issue 4, April – 2011.

[8] Shiva Nadi, Mohammad. H. Saraee, Mohammad Davarpanah Jaziand Ayoub Bagheri," FARS: *Fuzzy Ant based Recommender System for Web Users* ", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 1, January 2011 ISSN (Online): 1694-0814

[9] V. Seshadri, Sholom M. Weiss and Raguram Sasisekharan, "Feature *Extraction for Massive Data Mining*", From: KDD-95 Proceedings. Copyright © 1995, AAI

[10] Ms. Anjali Ganesh Jivani, "*A Comparative Study of Stemming Algorithms*", Anjali Ganesh Jivani et al, Int. J. Comp. Tech. Appl., Vol 2 (6), 1930-1938

[11] M. Pechenizkiy, S. Puuronen, A. Tsymbal, "FEATURE *EXTRACTION FOR CLASSIFICATION IN THE DATA MINING PROCESS*", International Journal "Information Theories & Applications" Vol.10

[12] Jasdeep Singh Malik, Prachi Goyal, [3]Mr. Akhilesh K Sharma [3]Assistant Professor, IES-IPS Academy, Rajendra Nagar Indore – 452012, India, " *A Comprehensive Approach Towards Data Preprocessing Techniques & Association Rules* "

