Rating Faculty for Foreign Student in Egypt by Bloom Filter Classifier and Collaborative Filtering

Mahmood A. Mahmood, Tarek A. Mohamed

Abstract: ***This paper presents an approach of bloom filter classifier and collaborative filtering to help foreign student to choose the suitable faculty according to his nationality and number of years that need to study. Our approach consist of three phases are: input phase, classification phase, and recommendation phase. In Input phase, the student enters the nationality and number of suggested years study. In classification phase, the approach classifies the student according to input data based on bloom filter classifier. In recommendation phase, the approach recommended the top five faculty if exists based on collaborative filtering technique (CF). Our dataset collected from Misr University for Science and Technology (MUST) and the results of our approach suitable and has a good manner for the student with accuracy 90%.***

Keywords: Recommender Systems, Collaborative Filtering, Bloom Filter Classifier, Rating Technique.

I. INTRODUCTION

In MUST there exist a problem for foreign student where the foreign student registered to MUST faculty according to his secondary school mark and the coordinate office only. The student has no experience about the courses and the knowledge of that faculty, so recommender system can help the student that don't know any knowledge about their faculty. Recommender system (RS) define as an intelligent system that provides advice to the user about a specific item aiding him in the decision-making process [1]. Such a system can gain information explicitly (user rating) or implicitly (user behavior). The recommender systems have traditional approaches are the follows:

A. Collaborative Filtering:

It is the simplest approached implements for RS applications, the basic idea depends on mathematical equations that compute the similarity between user-user or item-item and find the correlation between each other's, so this way called the user similarity or item-similarity [2].

B. Content-based:

It is an approach for RS, which uses the analysis of user profiles for rated previous items; or items that users visit it more times, so uses similarity measures to find similar or related items for that user [3].

C. Demographic:

demographic data are personal information data such as age, gender, location, etc..., this data used as inputs for RS application, RS divide data into some categories, where some users collects together for same age; or location; or gender; or hybrid of demographic data, so ge, gender, location, etc.) as input for recommendation, this approach combines groups that contain similarity users based on similarity demographic data, so collect the interests of each user in category and measure the similarity to find the similar items for each group [4].

D. Knowledge:

This approach uses artificial intelligent and machine learning to recommend items to users [2].

E. Hybrid:

This approach combines between different approaches in order to achieve the best performance or provides quality of recommendation. Recommender systems applied in many fields such as real live world (drinking water, climate change, grounded-level ozone prediction) [5], [6], [7], data mining [8], big data [9], [10], and location based social network [11].

Classification is a technique that used to classify the data into the specific label, classification categorized into single label and multi label. In single Label classification there exist only one decision attribute but in multi label classification there exist more than one decision attribute. Bloom Filter classifier is one of more classifiers, the heart of this classifier uses the principals of set theory in classification process. The rest of this paper is organized as follows. Section II, introduces basic concepts of bloom filter classifier and CF technique. Section III presents the materials and methods that we used in our approach. Section IV discuss the experimental results and presents the statistical analysis of the obtained experimental results. Finally, conclude the results of our approach

II. BASIC CONCEPTS

A. Bloom Filter Classifier

Bloom filter (BF) [12], valid spatial probability data structure, is used to represent a set of S elements It consists of a set. of bits m, denoted by BF [1, 2,... , M], Each set is initially zero. Describe elements in the set, the filter uses k independent hash functions (h1, h2 and... hk) Assuming that this hash ranges from 1 to m.

Revised Manuscript Received on May 21, 2020.

***** Correspondence Author

Mahmood A.Mahmood*, Department of Computer Science, Jouf University, Tubarjal, Kingdom of Saudi Arabia, Email: mamahmood@ju.edu.sa, Department of Information System and Technology, Cairo University, Faculty of Graduate Studies for Statistical Research (FGSSR), Email: mahmoodissr@cu.edu.eg.

Tarek A. Mohamed, Department of Information System

Misr University for Science and Technology Faculty of Business, Emai: tarek.mohamed@must.edu.eg.

These hash functions independently map each element of the universe to random number uniformly over the range. Each element x2S; BF [hi (x)] are set to 1 for 1¡ i ¡ k. Given Check membership by checking component y, bf. Bit at position hl(y); h2(y);... , hk(y) are set to 1. Each hi(y) (1 ¡i ¡k) are set to 1 and is taken into consideration Part S If not, you are not a member of group S. The accuracy of a bloom filter depends on the size of the filter m. Number of hash functions k, number of elements n. False positive can be computed by equation (1).

$$fp = (1 - e^{-kn/m})^k \quad (1)$$

B. Collaborative Filtering Technique

Collaborative Filtering (CF) Collaborative filtering is a technique that can select items that a user can be shared with similar users. It works by searching a large group of people and finding a smaller set of users with tastes like a user [13]. For a user-item pair (u, I) the prediction is composed of a weighted sum of the user ratings for the item most like I. The general item-based prediction algorithm has been formalized in the following equation (2).

$$pred(u,i) = \frac{\sum_{j \in ratedItems(u)} itemSim(i,j).r_{ui}}{\sum_{j \in ratedItems(u)} |itemSim(i,j)|} \quad (2)$$

There exist several variations for calculating the similarity between two items as shown in equation (3). The Person Correlation Coefficient alias Adjusted-Cosine Similarity [13], [14] is considered the most accurate similarity metric among all.

$$itemSim(i,j) = \frac{\sum_{u \in rated(i,j)} (r_{ui} - \overline{r}_i)(r_{uj} - \overline{r}_j)}{\sqrt{\sum_{u \subset rated(i,j)} (r_{ui} - \overline{r}_i)^2} \sqrt{\sum_{u \subset rated(i,j)} (r_{uj} - \overline{r}_j)^2}} \quad (3)$$

III. PROPOSED APPROACH

Proposed approach has three phases are (1) input phase, (2) Bloom filter phase, (3) Recommendation phase.

A. Input phase

In this phase, our approach cleaning data by select the best feature selection according to the graduate student reviews, and ask new student to enter his/her nationality and expected number of resident year in Egypt according to the number of year of faculty that hope to join it ; or enter the faculty that hope to join it. the input phase received the data and if student enter the name of faculty that hope to join the approach send the number of year of this faculty to the field of expected number of resident year in Egypt and ignore the hope faculty to student.

B. Bloom Filter Phase

In this phase, the data received from input phase that consist of two tuples. The first tuple consists of the feature selections of all graduate student reviewers, and the second consist of the information of new student. This phase applied the equation (1) to match the nationality of the student and get all graduate student that has the same nationality, after that, take the expected number of resident years in Egypt and scope to the all graduate students reviewers that has the same resident number of years.

C. Recommendation Phase

In this phase, the data received from Bloom filter phase as two tuples. The first tuple consists of information about the graduate student reviewers have the same nationality, and the second tuple has information about the graduate students that resident in Egypt the same expected number for new student. In this phase the approach uses equations (2,3) to rate the information and get the top-N faculties that enable to the new student to join to it according to the reviewers of graduate students.

IV. EXPERIMENTAL ANALYSIS AND DISCUSSION

A. Dataset and its Characteristics

The dataset collected from MUST university in Egypt, that dataset consist of the reviewing of 7923 graduate students in 14 faculties, the review consists of nationality, faculty, and number of resident years before graduation.

B. Experimental Results and Discussion

In the experiment divided the dataset into two part (80-20). First part contains 80% for training the approach and the remaining data considered as tested data. Also, dataset divided into (70-30) where 70% assumed as training set and 30% as test set. Finally, dataset divided into (50-50) and the result shown in table1 as shown in table1 the average accuracy is 90%.

Table 1: Average Accuracy For Dataset

Experiemnt size	#Test set	#training set	Accuracy
80-20	1585	6338	87%
70-30	2377	5546	89%
50-50	3961	3962	94%

C. Performance Evaluation

The accuracy computed as the following equation (4), average accuracy computed as the total number of all accuracies divided the number of accuracies.

$$accuracy = \frac{numberofpositivereviewersstudents}{allnumberofmatchedreviewersstudent} \quad (4)$$

V. CONCLUSION

This paper has introduced the challenge of the force of foreign student to choose the faculty that can be completed faster than another faculty that can be introduced to it. First, we assume that the student has many choices of more than one faculty. Moreover, we proposed a proposed approach to recommend to student the rating of faculties that can be introduced to it before entering the university. In addition, we considered the nationality and number of resident years for student is the selected features for student and the result has a good manner for new student depending on the reviewing of graduate students with accuracy 90%. In the future, we increase the features selection for student such as the high grades (more than 75%) of secondary school courses to enhance the accuracy of our approach.

REFERENCES

1. MA. Mahmood,N El-Bendary, J Platos, AE Hassanien, Ha Hefny, an intelligent multi-agent recommender system, Innovation in bio-inspired computing and applications, pp.201-213,2014.
2. Mahmood M.A., El-Bendary N., Platos J., Hassanien A.E., Hefny H.A.(2014) An Intelligent Multi-agent Recommender System. In: Abraham
3. A., Kromer P., Snasel V. (eds) Innovations in Bio-inspired Computing and Applications. Advances in Intelligent Systems and Computing, vol 237. Springer, Cham
4. P.Lops, M.de Gemmis, G.Semeraro, Content-based Recommender Systems: State of the Art and Trends, in: F.Ricci, et al. (Ed.), Recommender Systems Handbook, pp. 73-106, 2011.
5. L.Safoury, A.Salah, Exploiting User Demographic Attributes for Solving Cold-Start Problem in Recommender System, Lecture Notes on Software Engineering, Vol. 1, No. 3, pp.303-307, August 2013.
6. S.Mahmoud, N.EI-Bendary, M. Mahmood, A.EllaHassanien,An Intelligent Recommender System for Drinking Water Quality ,Hybrid
7. Intelligent Systems (HIS), 2013 13th International Conference on.IEEE, 2013.
8. MA Salam,MA Mahmood,YM Awad,M Hazman, N EL Bendary,Aboul Ella Hassanien, Mohamed F Tolba, Samir Mahmoud Saleh, "Climate recommender system for wheat cultivation in North Egyptian Sinai Peninsula", proceedings of the fifth international conference on innovations in bio-inspired computing and applications IBICA 2014, pp.121-130, Springer international publisher, 2014.
9. M.Mahmood, E.Tamah Al-Shammari, N.El-Bendary, A.EllaHassanien, H.Hefny, Recommender System for Ground-Level Ozone Predictions in Kuwait, Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on. IEEE, 2013.
10. U.Panniello, A.Tuzhilin, M. Gorgoglione, C.Palmisano, A.Pedone, Experimental Comparison of Pre- vs. Post-Filtering Approaches in Context-Aware Recommender Systems, Proceedings of the third ACM conference on Recommender systems.ACM, 2009.
11. M.Lee, A Recommendation System based on Big Data Technology, International Journal of Science and Applied Information Technology, Vol 3, No.5, September - October 2014.
12. X.Han, L.Tian, M.Yoon, M.Lee, A Big Data Model supporting Information Recommendation in Social Networks, Cloud and Green Computing (CGC), 2012 Second International Conference on.IEEE, 2012.
13. Symeonidis, Panagiotis, DimitriosNtempos, YannisManolopoulos. Recommender Systems for Location-based Social Networks.Springer, 2014.
14. Bloom, B.H., "Space,time trade-offs in hash coding with allowable errors",Commun, ACM 13(7), 422-426,1970
15. Dan Frankowski Jon Herlocker Schafer, J. Ben and Shilad Sen. Collaborative filtering recommender systems. In Brusilovsky, Peter, Alfred
16. Kobsa And Wolfgang Nejdl (Editors): The Adaptive Web, 4321:291-324, 2007.
17. Brent Smith Linden, Greg and Jeremy York. Amazon.com recommendations:Item-to-item collaborative filtering. Ieee Internet Computing, 7:76-80, 2003.

AUTHORS PROFILE

MAHMOOD A. MAHMOOD, completed Ph.D. degree in June 2014 in Computer Science and Information from Cairo University, Institute of Statistical Studies and Research (ISSR). I have received my B.Sc., 1998 from Faculty of Science, Cairo University and Master of Computer Science and Information in 2009 from Statistical Studies and Research, Cairo University. Now, I am a Assistant Professor in the ISSR, Cairo University, Egypt. I have 17 years, experience in Teaching at Cairo University, Egypt. Currently, he is working as an assistant professor at Jouf University, Tubarjal, Saudi Arabia. My research interests include Artificial Intelligent, Data Mining, Recommender Systems and Multimedia.

Tarek A. Mohamed, completed Ph.D. degree in June 2014 in Computer Science and Information from Cairo University, Institute of Statistical Studies and Research (ISSR). I am a Assistant Professor Department of Information System Misr University for Science and Technology Faculty of Business.