ISSN (Print): 0974-6846 ISSN (Online): 0974-5645

# A Survey on Collaborative Categorization using Fuzzy Logic for Improved User Suggestions

M. Anjani and V. Vijaya Kaveri

Sathyabama University, Chennai-600119, Tamil Nadu, India; Anjani.m2008@gmail.com, vijirama.kaveri@gmail.com

#### **Abstract**

Background/Objectives: This paper presents a review on the usage of social tags that can be employed for recommending in social networks. It also focusses on pros and some algorithms as solutions to the problems. Recommender systems are the one used for providing information regarding the categories that belongs to the items which makes use of fuzzy clustering, in which the membership degree of the object is not available in the calculated categories. If calculated categories are available it will provide good quality recommendations to the user. Methods/Analysis: The method suggested in this work is to utilize fuzzy logic to classify the similar users from the clustered users based on their profiles and taste of viewing similar items. Fuzzy logic here plays a vital role in providing recommendation to the similar users once the classification is done. Findings: Hence, by surveying some papers on recommendation in social networking, we propose a solution by creating a dynamic system which would provision recommendations to the user to make the domain selection more effectively. The domains refers to books, movies, songs, which are purely dynamic. This methodology helps to display the most viewed item as the top most item and displays the item ratings in the webpage. The main approach used here is fuzzy logic with the detailed representation of features of objects and modeling user profiles. Applications/Improvements: This type of dynamic domain system will provide better results when compared to the conventional recommendation systems as it allows users to give comments on various products on different domains. This type of recommendation can be consumed by many social networking sites for gaining page rating and it can be further improved in future by combining the online dictionary to classify the terms used in user rating.

Keywords: Collaborative Filtering, Fuzzy Logic, Recommender Systems, Simmilarity, Tags, User Profiles

## 1. Introduction

The web generally means providing huge amount of data from various sources to the users with different and similar tastes. The web may be an information oriented webpage, e-commerce sites, social networking sites. Recommender systems plays a vital role in this context. Recommender systems are the one that provides the recommendation for a site or a particular item content in that site. The recommendation to these sites are generally provided to the users by using the recommendation provided by other users. These recommendation systems works based on the user interest. Initially the recommender systems monitor the user interest by using some of the techniques like collaborative filtering, content based algorithms

and fuzzy clustering. In this paper, a recommender system with a dynamic framework is being concentrated. Collaborative filtering is employed for collecting the user profiles from the users. It compares the users' view and finds the similarity between the users to find their likes and dislikes. Based on the user preferences their tastes are determined. CF integrates time information along with the content information for providing recommendation. CB is a technique that consists of sequential steps for finding whether the item is being liked by the user or not. It compares the item content between the users for further processes in this type of system. The item here refers to a set of descriptors or terms. The terms are being assigned either automatically or manually. Fuzzy clustering is used in datamining for grouping similar objects which has

<sup>\*</sup> Author for correspondence

a drawback of finding the standard result<sup>1</sup>. However in recommendation systems it is meant for grouping users based on their profiles and item content. This paper provides a review on these techniques and provide a novel approach in using a recommender system.

## 2. A Survey on Recommender Systems

The recommendation to the users are given using various approaches followed in recommender system and a concept called social tagging is used. The information generated by the tagging systems can be utilized for recommendation purposes. Social tagging has improved the social and personal experience of users across the web by enabling the task of organizing, managing, sharing and searching the data in web. However, if some unrelated tags that is created by a user which produces noise and the tags can be unreliable; thus, making recommendation a nontrivial task. In this study, a recommender system is proposed based on the similarities between user and item profiles along with a time tag. The approach applied is to generate user and item profiles by identifying tag patterns that are frequently generated by users. These tag patterns are categorized into irrelevant patterns and relevant patterns. This study focusses on various problems in recommender systems and an approach to improve the performance of the recommender system.

#### 2.1 Linguistic Quantifiers

The natural human language present in a tag may contain the same information or different information. These information may be represented in different ways. For an instance, a tag information of many tags denoting the same information may contain different words to provide the same meaning. To avoid the ambiguity problem, the tags that contain same information should be recognized. The linguistic quantifier generally quantifies the quantity and used in logics for constructing new formula from the existing ones. It can be considered to improve the performance of the system<sup>2</sup>. It can be utilized in recommender systems in a way that the statements has to be analyzed in order to find out the semantics and summarizing the statement. It is expected that a statement satisfies most or all the criterion<sup>3</sup>. The weights satisfies the following condition,

$$\mathbf{w}_{i} = (\mathbf{i} | \mathbf{n}) - ((\mathbf{i} - \mathbf{n}) | \mathbf{n}) \tag{1}$$

This works quite similar to a thesaurus thereby analyzing the meaning of the words to summarize.

#### 2.2 Sparsity

Sometimes a huge dataset or database generally contains sparse data values. It be defined as the high percentage of variable's cell that do not contain actual data. This means that the empty values will occupy the memory space in the file. There are two types of sparsity. First, controlled sparsity that contains a range of values for one or more dimensions with no data. Second, random sparsity is the one in which null values are scattered throughout the data. Sparsity is one of the challenges in CF systems<sup>4</sup>. However there are variety of solutions provided for this problem by combining the locally similar neighbors with globally similar neighbors. This can be given as,

$$Pred_{R} = (1-\alpha)^* predR_{I} + \alpha^* predR_{G}$$
 (2)

where, RL- predicted rating through local neighbor and RG- predicted rating through global neighbors. αvariation of local and global similarities.

On taking into account not only the overall rating sparsity is considered instead the user-level item is also considered. An automatic weighting scheme is being used through evolutionary approach to obtain unified measure of sparsity. There are so many techniques used to solve this sparsity problem. Model-based techniques like svd, latent semantic indexing techniques can be utilized to avoid this problem. Hybridization of collaborative filtering along with the other filtering techniques also solves data sparsity problem. Other solutions covers the rules generated by apriori algorithm and the recursive algorithm allowing the nearest neighbors to join the prediction process.

## 2.3 Collaborative Filtering

There are so many approaches used in recommender systems. One of the approaches is collaborative filtering. It finds the user preferences by identifying the likes and dislikes of the user. It extracts the user profiles as a tags along with their item interests to categorize them as similar or like-minded users<sup>5</sup>. The user similarity can be computed using Pearson correlation coefficient,

$$p_{a,u} = \frac{\sum_{i=1}^{m} (r_{a,i} - \overline{r_a}) * (r_{u,i} - \overline{r_u})}{\sqrt{\sum_{i=1}^{m} (r_{a,i} - \overline{r_a}) * (r_{u,i} - \overline{r_u})}}$$
(3)

ra, i is the rating given to item i by user a and rais mean rating by user a. pa,u is the similarity between the user a and u.

This considers the user ratings for computing the resource recommendation. Traditionally, log based recommendations are based on binary values 0 or 1 denoting whether a user viewed or transacted a resource<sup>6</sup>. But there is a problem with this approach that it provides static recommendation lists as the users' behaviors may change with respect to time. For example, at one point the system would have calculated the users' interest for providing recommendation later. But the users' interest might get shifted as time goes. This can be approached in a way by providing a solution by constructing a modified rating matrix along with the time information or tag information or the addition of both. For this matrix construction any one of the two approaches are followed. User- resource binary matrix is based on traditional log based matrix in which the value is 1 if the user bookmarks a resource otherwise 0. Second is modified user rating matrix which can contain either time or tag or combination of both information<sup>6</sup>. Tag weight strategy indicates the user preferred resource whereas the time weight determines at what time the user has bookmarked the resource and given as an equation,

$$W_{tag}(u,r) = \sum_{ta \in tag(u,r)} W_{u,ta}$$
 (4)

Wtag(u,r) denotes the tag weight, tag(u,r) are the set of tags bookmarked. U, ta is the tag score of each tag.

$$W_{time(u,r)} = \exp\{-\ln 2^* time(u,r) | hlu\}$$
 (5)

Wtime(u,r) denotes the time weight, time(u,r) is the time the resource is bookmarked and non-negative integer. Hlu is the half- life of each user.

$$M_{u,r} = \lambda W_{tag}(u,r) + (1-\lambda) W_{time}(u,r)$$
 (6)

Mu,r is the time tag information and  $\lambda$  is to adjust the significance of both the weights.

#### 2.4 Fuzzy Logic

Fuzzy sets were introduced by Zadeh. It does not deal with the crisp set. These logics are usually used in controller systems. It can be used for determining actual value from the range between 0 through 1. This more over resembles a human decision unlike the other systems. Appropriate distance function aids to match different users with fuzzy model. Fuzzification is a concept involved in this approach to group the users based on the distance measures. Two fuzzy distance methods are generally used. One is the use of IF-THEN rules and the other one is finding out the fuzzy distance between the users<sup>7</sup>. According to fuzzy concepts local distance measure is preferred rather than preferring global distance because it works well for minimum distances and discards the remaining<sup>7</sup>. The distance function for a and b can be given as,

$$fd(a,b)=d(a,b)*d(a,b)$$
(7)

Where d(a,b) is difference operator, a and b are vectors of size l. d(a,b) is of any vector distance metric which can be computed using Euclidean distance formula<sup>7</sup>. Therefore fuzzy sets are employed in control systems for decision making process.

## 3. Conclusion

This survey presents an idea on recommender systems and the major approaches involved; collaborative filtering, fuzzy logics and linguistic quantifiers along with the problems and solutions. The recommender systems can be made more effective by a novel approach by creating a dynamic framework rather than considering a single dataset or resource (that is, the recommender system with multiple domains and categorization of domains based on the item content for generating tags) and the time tag can be provided along with the tag information. This can help to provide a better recommendation to the user.

## 4. Acknowledgement

The authors gratefully acknowledge the encouragement and support of Sathyabama University, Chennai in the publication of this paper.

## 5. References

- 1. Revathy S, Parvaathavarthini B, Rajathi S. Futuristic validation method for rough fuzzy clustering. Indian Journal of Science and Technology. 2015 Jan; 8(2):120–27. DOI: 10.17485/ijst/2015/v8i2/58943.
- 2. Zenebe A, Norcio AF. Representation, similarity measures

- and aggregation methods using fuzzy sets for content-based recommender systems. Fuzzy Sets and Systems. 2009 Jan; 160(1):76–94.
- 3. Anand D, Mampilli BS. Folksonomy-based fuzzy user profiling for improved recommendations. Expert Systems with Applications. 2014 Apr; 41(5):2424–36.
- 4. Anand D, Bharadwaj KK. Utilizing various sparsity measures for enhancing accuracy of collaborative recommender systems based on local and global similarities. Expert Systems with Applications. 2011 May; 38(5):5101–09.
- 5. Babu MSP, Kumar BRS. An implementation of the us-

- er-based collaborative filtering algorithm. International Journal of Computer Science and Information Technologies. 2011; 2(3):1283–86.
- 6. Zheng N, Li Q. A recommender system based on tag and time information for social tagging systems. Expert Systems with Applications. 2011 Apr; 38(4):4575–87.
- 7. Al-Shamri MYH, Bharadwaj KK. Fuzzy-genetic approach to recommender systems based on a novel hybrid user model. Expert Systems with Applications. 2008 Oct; 35(3):1386–99.