International Journal For Technological Research In Engineering
Volume 3, Issue 10, June-2016 *ISSN (Online): 2347 - 4718*

SIMILARITY BASED RECIPE RECOMMENDER

Vivek MB
Information Science and Engineering, Sri Jayachamarajendra College of Engineering, Mysuru, India.

Abstract: Recommender systems are becoming an important part of any internet user. It is actively involved in finding relevant information. The users become an important part of the recommender system. In this paper, we present two of the approaches to recommend recipes based on user preferences for the application dataset. The two approaches used are the item based collaborative filtering approach and user based collaborative filtering approach. The similarity techniques used are the Log Likelihood and Tanimoto for item based and Pearson Correlation and Euclidean Distance for the user based approach respectively. Good recommendations are observed in case of the user based collaborative filtering approach than the item based collaborative filtering approach.
Keywords: Recommender Systems; Item based collaborative filtering; User based collaborative filtering.

I. INTRODUCTION

Recommender systems are commonly used in e-commerce websites to recommend related items to the users using the websites. Recommender systems produce an ordered list of recommendation as per the user way of using the website. The main goal of a recommender system would be to provide the user with appropriate recommendations. Recommendation for movies on Netflix, followers on Twitter are real world examples of the way the recommender system are used. The design of a recommender system is domain specific. Recommender systems can be built for various domains making use of the content based approach or the collaborative filtering approach [1]. Even though it is domain specific the main thing is interactions of the users with the application or website. Collaborative filtering finds the similar users according to the preferences but content based approach finds the similar items based on the content present in the items. In this paper, we make use of collaborative filtering approach to make good recommendations for the users based on preferences in the form of ratings. The remainder of this paper is organized as follows. In Section 2 we give a brief description of related work. In Section 3, we discuss the methodology. In Section 4, the experiments and results are discussed. Conclusion is discussed in Section 5.

II. RELATED WORKS

Recommender systems are built for various domains like twitter followers [2], e-commerce [3] etc. Recommendations are made using either the collaborative filtering approach or the content based approach [4]. Collaborative filtering identifies the patterns of the different users and identifies similarities between them. In content based it is more of identifying the content present in the recommended items [5]. Both approaches can also be combined to make a hybrid recommender system [6]. To our knowledge there are relatively a very few attempts in the field of recipe recommendation based on the user preferences.

III. METHODOLOGY

Our application recommends recipes by making use of the Collaborative Filtering approach. Initially recommendations for various recipes are based on the main ingredients. The user selects the ingredients and according recommendations are made. Then if a user is interested in the initial set of recommendations, he would rate the recipes on the scale of five to one according his tastes. To identify the similar recipes we make use of Tanimoto Coefficient Similarity [7] [8] and Log Likelihood Similarity [9]. To identify similar users we make use of the Pearson Correaltion [10] and Euclidean Distance [11].

A. Item Based Recommendation

Here, Item based recommendations are based by identifying similar recipes based on the main ingredients and also it considers the ratings given by the user, even though it doesn't take into account the actual rating values. It considers only if a user has rated a recipe or not. It is based on recipe co-occurrences by the usres. The similarity values are used to obtain a ranked list of recommended recipes. To calculate the similarity, we make use of two similarity measures namely, Tanimoto Coefficient similarity and Log Likelihood similarity. Tanimoto Coefficient is an extended Jaccard coefficient. It is the number of recipes that two users express some preference for, divided by the number of recipes that either user expresses some preference for. The actual rating values do not matter, only their presence i.e. 1 or absence i.e. 0, does. When two of the recipes overlap, the result is 1.0. When they do not have anything in common, it is 0.0. The value is never negative.
Tanimoto coefficient [12] is given by:

$$T(x,y) = \frac{N_z}{N_x + N_y - N_z}$$

Where,
Nx - Number of customers who would rate item X
Ny - Number of customers who would rate item Y
Nz - Number of customers who would rate both items X&Y
Log-likelihood-based similarity [13] is similar to the Tanimoto coefficient based similarity. Even Log Likelihood doesn't take into account the value of the ratings. It is mainly based on the total number of recipes common between the two different users, but the value indicated is more of how unlikely it is for two unique users to be having so much overlap, given many recipes present and the number of recipes each of the user has rated for. To compute the score, let counts be the number of times the different events

occurred together (e_11), the number of times each of the events has occurred without the other event taking place (e_12 and e_21) and number of times neither of all of these events taken place (e_22). By having all these information Log-likelihood ratio would be computed as,
LLR = 2 sum(e) (H(e) – H (row Sums (e)) – H (col Sums(e)))

B. User Based Recommendation
User based recommendations are based on identifying how similar are two users by making use of the ratings given. To identify similar users we make use of Euclidean Distance and Pearson Correlation. The Euclidean Distance [14] would be computed as

$$\sqrt{n}/(1 + distance)$$

where n would be the number of dimensions. $\sqrt{n}$ is chosen because randomly-chosen points have a distance that will grow as $\sqrt{n}$.
Pearson correlation [15] for users X and Y is given as,
$\sum X^2$: sum of the square of all X's rating values.
$\sum Y^2$: sum of the square of all Y's rating values.
$\sum XY$: sum of the product of X and Y's rating value for all items for which both X and Y give a rating.
The correlation is then

$$\sum XY \Big/ \sqrt{\left(\sum X^2 * \sum Y^2\right)}$$

The correlation may be looked upon as the cosine of the angle between the two vectors identified by the users rating values.

IV. EXPERIMENTS AND RESULTS

For our work the recipe data as well as the user data in the form of ratings is obtained by making use of the developed application. There are about 20 users, 78 recipes with 208 user preferences. We implemented the item based collaborative filtering approach making use of the ratings given by the users' and using two similarity techniques namely Tanimoto Coefficient Similarity and Log Likelihood similarity. The performance of this approach is measured using the Recall measure based on main ingredients. The recall measure is defined as

$$Recall = \frac{|\{\text{ relevant recipes}\} \cap \{\text{retrieved recipes}\}|}{|\{\text{relevant recipes}\}|}$$

The results obtained for the recall measure was 2% using the Tanimoto similarity and 1% for Log Likelihood similarity. We implemented the user based collaborative filtering approach by making use of the ratings given by each of the users for the interested recipes. The results are in the form of ratings that would have been given by the users if a particular recipe would have been recommended. The performance is evaluated using the average absolute difference metric. It is given as

$$AAD = \frac{\sum_{i=0}^{n}(Difference\ between\ actual\ and\ estimated\ ratings)}{Total\ number\ of\ recipe\ rated}$$

The best results for our dataset were obtained when we used Euclidean Distance similarity metric. The difference between the actual ratings and the estimated ratings was about 1.1666 which is very reasonable in the case of making good recommendations. The performance was better by making use of the user based approach rather than the item based approach. Even though the running time was better in case of item based approach, better recommendations were made and also the performance was better in the case of user based approach.

V. CONCLUSION

In this paper, we implement two approaches for recommending recipes based on user ratings for the interested recipes. Recommendations are made quicker in item based collaborative approach. But appropriate recommendations are made in case of user based collaborative approach. Results of the user based collaborative approach are found to be better than the item based collaborative approach. This approach can be combined with a content based approach to obtain a hybrid approach and can be applied in various fields accordingly.

REFERENCES

[1] Adomavicius, G., Tuzhilin, A. "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," IEEE Transactions on Knowledge and Data Engineering, 17(6), pp. 734-749, 2005.
[2] Pankaj Gupta, AshishGoel, Jimmy Lin, Aneesh Sharma, Dong Wang, and Reza BosaghZadeh "WTF:The who-to-follow system at Twitter," Proceedings of the 22nd international conference on World Wide Web WWW 2013, Rio de Janeiro, Brazil.ACM 978-1-4503-2035- May 13–17, 2013.
[3] Dhohaalmazro, GhadeerShahatah and Lamia Albdulkarin, A survey paper on Recommender Systems, arXiv:1006.5278v4 [cs.IR] 24 Dec 2010.
[4] HoseinJafarkarimi; A.T.H. Sim and R. Saadatdoost "A Naïve Recommendation Model for Large Databases," International Journal of Information and Education Technology, June 2012.
[5] R. J. Mooney and L. Roy. "Content-based book recommendation using learning for text categorization. In Workshop Recom." Sys.:Algo. and Evaluation, 1999.
[6] Samanerajabi , Ali Harounabadi and Vaheaghazarian, "A recommender system for the web:Using user profiles and Machine Learning Methods," International journal of Computer Applications (0975-8887) Volume 96-No. 11, June 2014.
[7] Seifoddini, Hamid.Djassemi, Manocher. "Merits of the Production Volume Based Simlilarity Coefficient in Machine Cell Formation," 2007.
[8] Abreu, Rui. Zoeteweij, Peter.VanGemund,Arjan. "An Evaluation of Similarity Coefficients for Software Fault Localization," 2010.
[9] "Evaluating and Implementing Recommender Systems As Web Services Using Apache Mahout,"http://cslab1.bc.edu/~csacademics/pdf/14C

asinelli.pdf
[10] "Pearson Correlation: Definition and Easy Steps for Use," http://www.statisticshowto.com/what-is-the-pearson-correlation-coefficient/
[11] A. Madylova and S.G. Oguducu, "A taxonomy based semantic similarity of documents using the cosine measure," in Proceeding of International Symposium on Computer and Information Sciences, pp. 129–134, 2009.
[12] "Discussion of Similarity Metrics" http://mines.humanoriented.com/classes/2010/fall/csci568/portfolio_exports/sphilip/tani.html
[13] "Surprise and Coincidence - musings from the longtail,"http://tdunning.blogspot.in/2008/03/surprise-and-coincidence.html
[14] "EuclideanDistanceSimilarity,"https://builds.apache.org/job/MahoutQuality/javadoc/org/apache/mahout/cf/taste/impl/similarity/EuclideanDistanceSimilarity.html
[15] "PearsonCorrelationSimilarity,"https://builds.apache.org/job/MahoutQuality/javadoc/org/apache/mahout/cf/taste/impl/similarity/PearsonCorrelationSimilarity.html

