IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 4, Ver. I (Jul - Aug 2017), PP 14-19
www.iosrjournals.org

A Study on Ontology Based Collaborative Filtering Recommendation Algorithms in E-Commerce Applications

[1]H Mohana, [2]Dr. M Suriakala
[1]*Ph.D Research Scholar, Department of Computer Science, Dr Ambedkar Government Arts College (Autonomous), Vyasarpadi, Chennai*
[2]*Assistant Professor and Research Advisor, Department of Computer Science, Dr Ambedkar Government Arts College (Autonomous), Vyasarpadi, Chennai*

Abstract: *Recommender system is a growing proliferation in today online applications contributed to the problems of Information overloading. In a day to day life enormous amount of data is generated and collected leads to a problem of information overloading. This paper focuses on how to deal with the problem of information overloading and how to recommend an additional product to the end user using collaborative filtering (CF) recommendation algorithms. The personalized recommendation algorithm with their benefits and limitations are described. A pitfall occurs in CF recommendation system is described. An outline framework is proposed for the initial stage of recommendation. A sample problem statement is framed to show how recommendation technology is embedded in to the framework.*

Keywords: *Web Mining, Collaborative filtering, CF Algorithms, CF Framework, E-Commerce*

I. Introduction

Electronic Commerce / E-Commerce is an iconic representation for online commercial transactions that allows business applications to deal with different organizations. In an internet world E- commerce industry plays a vital role in the online digital transactions. E-commerce websites becomes an unbeatable one and it shows the rapid development in it. The effectual technology used to extract information from E-Commerce website is data mining and web mining. Data mining is the process of extracting immersing patterns from large database. Whereas web mining is the usage of data mining techniques to extract interesting information from the web data. E-Commerce firm such as Amazon, Alibaba, Snap deal enlarge the plan of upscale, Customers easily find items on recommender systems[1].

This recommender system uses information which is extracted using web mining techniques to recommend the related searches to the users of ecommerce sites. Enterprises mainly focus on rebuild the relation with their old customer and simultaneously focusing on new customer by applying recommender systems. "The main goal of recommender system is to generate meaningful recommendations to a collection of users for items / products that might interest them"[2]. The use of recommender system in an ecommerce environment can impact the end users as well as merchants to achieve both of their goals. Therefore, in an E-Commerce industry a recommender system acts as an interface between buyer and the seller.

Recommender system plays a vital role in filtering the information in several E-Commerce application areas like movies, music, news, books, research articles, search queries, products / items etc. since there is enormous amount of data store on these digital commercial transactions, it is difficult for a customer to make his search on relevant products or items to compare and a seller to recommend a related key search to the end user. Recommender system supports to sort out these issues and helps the customers to increase their accuracy level of getting the relevant and related search. So, that to overcome this issue a Recommender system was introduced. The following figure represents the taxonomy of recommender system in a hierarchical tree like structure.

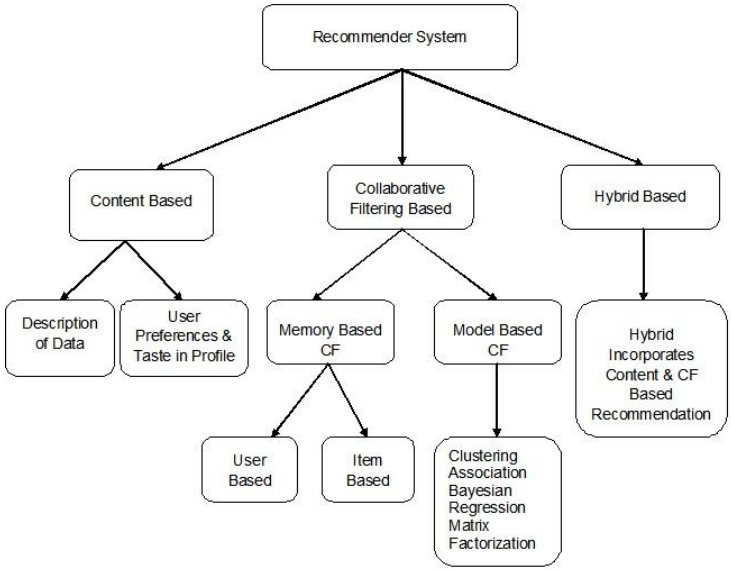

Fig (1): Taxonomy of Recommender Systems

The taxonomy explains that the recommender system is classified as three different approaches and they are Content based approach, Collaborative filtering (CF) based approach, and Hybrid based approach. CF approach is classified in to two type's memory based and model based approach. Memory based approach is again classified on the attributes of user based and item based approach. Similarly, Model based approach is classified on the basis of attributes which are processed using the techniques of clustering, association, regression, Bayesian, matrix factorization. Whereas, hybrid incorporates content based and CF based recommendation types.

1. Content Based Recommendation Technique

It is also called as cognitive filtering and recommendation is purely based on Content or description of the product/item/key term searched by the active user. User profile that holds the Demographic data as well as information about user's tastes, preference and needs are evaluated. User similarity index is calculated taking in to the account of similar features of these users. There is no need to worry about the user's purchasing history, rating history etc.

Limitations

The accuracy for the recommender system grows significantly down, when the system recommends the user based on any single context. So, it is difficult to recommend the user's Preference depends on the content. For example, if a user search for a product Samsung 4G phone the search content will be like this "Samsung 4G phone". The recommender system will provide details only on the features of the search about Samsung 4G phone, instead of providing details with same features of different brands like Apple iPhone 7s, Redmi note5 etc.

2. Collaborative filtering recommendation techniques

CF approach is purely based on user's browsing history, purchasing history and rating history. User similarity index is calculated depends upon the similar features present in some other user's behavior in searching their product and it is considered as one of the most prevailing and promising approaches in recommender systems. Prediction is made on what the users will like based on the similarity with other users. CF approach makes recommendation according to the nearest neighbor preferences. This approach is again classified as two types Memory Based CF approach and Model Based CF approach.

A. Memory based CF approach

Memory based CF approach calculates the similarity preferences between users according to the existing data sets and selects the users who has high similarity index will become the nearest neighbor to the target user. Memory based CF approach is categorized in to two types User based CF and Item based CF and

these approaches uses the nearest neighbor algorithm to recommend the nearest products to the users search. Memory-based methods employ the entire user-item database to generate a prediction[3-5].

I. Traditional Memory Based CF Recommendation Algorithms

MEMORY BASED CF RECOMMENDATION		
ALGORITHMS	BENEFITS	LIMITATIONS
• Neighbor based CF • User based nearest neighbor algorithm • Item based nearest neighbor algorithm • Top N recommendation • User based Top N recommendation algorithm • Item based Top N recommendation algorithm	• Easy implementation • New data can be added easily and incrementally • Need not consider the content of the items being recommended • Scale well with co-rated items	• Depends on user ratings • Performance decreases when data are sparse. • Skewing occurs when the no of users who has rated a common item. • Cannot recommend for a new users. • Scalability problem.

Table 1

B. Model Based CF Approach

A model based approach is generated by learning the training data sets and the trained dataset will re – trained periodically as the models are updated. It is also called as "Model – Learning" phase. A typical model based CF technique includes Matrix factorization, Association, clustering, Bayesian, Regression, Markov decision process (MDP), etc. Model-based method use demographic and content based information to create a model that generates recommendations[5-9]. The proposed model is used to make predictions at run time.

II. Traditional Model Based CF Recommendation Algorithms

MODEL BASED CF RECOMMENDATION		
ALGORITHMS	BENEFITS	LIMITATIONS
• Bayesian belief network CF algorithm. • Clustering CF algorithm • Regression based CF algorithm. • MDP based CF algorithm.(Markov decision process)	• Improves prediction performance. • Gives better result for sparsity and scalability problem. • It gives high classification accuracy for both complete and incomplete data. • It gives good prediction for numerical values.	• Expensive model building. • Trade off between prediction, performance and scalability. • Inability to deal with incomplete data. • Lose useful information for dimensionally reduction technique.

Table 2

CF Limitations

- **Cold Start Problem**
  It is one of the most exigent problems in recommender systems. Recommendation is a quite difficult task for a novice user or new user. In case of new users, the system doesn't have enough information about their preferences in favor of to make recommendations. A new item which is going to launch in an ecommerce era won't have prior rating. Therefore, when there is a new user that has been registered to the system and no prior rating of this item is found in the rating table[10] then a cold start problem occurs.
- **Data sparsity**
  Sparsity occurs when the number of ratings needed for prediction is greater than the number of ratings obtained because CF approach requires user explicit expression of personal preferences for products[11]. Some users who do not always rate the items purchased by them. This leads to the problem of data sparsity.
- **Scalability**
  When the data are sparse their scalability problem occurs. Scalability problem is caused in order to process large amount of information. The scalability problem is related to performance problems in the search for neighbors in memory based algorithms[11].
- **Poor security**

3. Hybrid Approaches

A variety of techniques has been proposed for performing recommendation, including content-based, collaborative based. To improve better and accurate performance, these methods are combined to give hybrid recommendation. It is a combination of content based and the CF based recommendation techniques. The techniques which combine the multiple recommendation approach. The hybrid approach deals with the problem of conventional recommendation systems like Netflix, Google News Recommendation System, etc

III. Traditional Hybrid Based CF Recommendation Algorithms

HYBRID BASED CF RECOMMENDATION		
ALGORITHMS	BENEFITS	LIMITATIONS
• Hybrid combines content and CF algorithms.	• Overcomes content and CF algorithm limitations. • Better performance. • Improves prediction performance.	• Implementation is expensive. • Increased complexity. • Needs external information that usually not available.

Table 3

Ii Related Work

Alfredo cutolo et.a[l2] recommends a product and services through e-service semantic description. A structural characteristics of product/ service is defined 1) User characteristics 2) user behavior vector space model approach exploits a weighted score to represents user profit, user action, service and context elements his future work is to handle with real and synthetic data.

Wu sen et.al[13] integrated CF recommender system with interest drift based on forgetting function change of user Interest will degrade the accuracy. He implemented an exponential function to redefine the weight of each item. Traditional user similarity index result is compared with his proposed improvised algorithm. Movielense data set is taken and show proved that his proposed work is better than the traditional one.

Zhe Yang et.al[14] Studied and utilized how to predict the user interest on mobile communication and how to make a proper recommendation. He proposed a framework for CF recommender system. Two case studies are described to validate his proposed framework.

Chen Li et.all[15] Chosen a journalism industry for recommendation. His research work mainly focuses on to propose a new personalized recommendation algorithm for news websites based on user's browsing path. *Markov chain* model is used to identify the user's browsing path user similarity index is calculated using improvised Jaccard similarity co-efficient formula.

Fuli Zhan[16] *proposed* a book recommendation algorithm based on time sequence and space distance for college and university libraries. His also targets two important characteristics Time sequence to borrow books and *Book* rotation time.

Bushra Alhisawi et.al[17] concentrates on how to compare the user similarity metric. Therefore, a new genetic algorithm is designed called SimGen. He finally compares his SimGen also with traditional similarity algorithm called Pearson correction, vector cosine based etc. He proved that SimGen shows high accuracy that other related algorithms.

Zhipeng Zhang et.al[18] Targets a user based CF for selecting the nearest Neighbor. In order to remove the redundant user from all other user. He followed two algorithms one is Traditional UBCF algorithm another is covering-based CF approach. His Future work is to deal with cold-star problem.

III A common framework for CF recommender system:

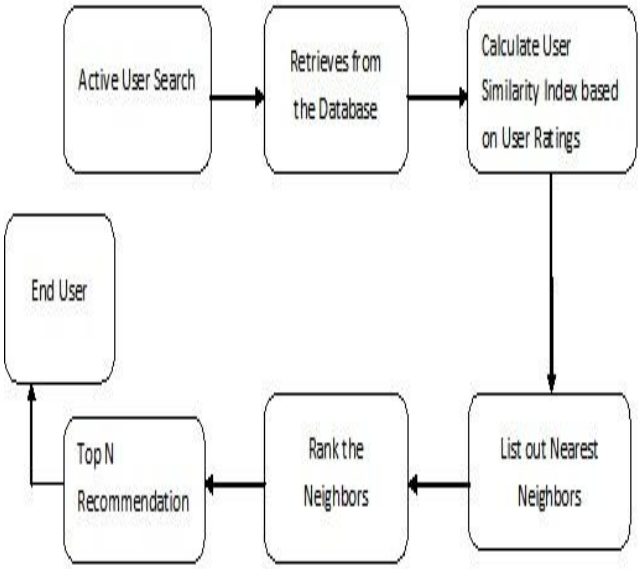

Fig (2): CF Framework Structure

IV How Recommendation Approaches Works in A Common Framework

Problem Statement:

Consider that a set of additional products needs to be recommended to a customer who purchased a product X.

A common framework describes that how CF recommender approach works well for the above problem statement. Product X enters in to the database, whose database will check for some other users who purchased a product X, bought what product next? User similarity index (USI) is calculated. There are several formulas which measures the user similarity index. The traditional formula to evaluate the user similarity index in collaborative filtering approach is Pearson correction coefficient method, cosine similarity measure, Jaccard similarity measure etc. The output of USI will list out the nearest neighbor and the collaboration filtering recommendation algorithm is applied to find out the Top-N from the list. Finally, the product which is relevant to the most will be recommended to the end user. As a whole the CF based recommendation algorithm predict the missing ratings, rank items and selects the Top-N from the list.

IV. Conclusion

This article discusses about an ontology based collaboration filtering recommendation algorithms for various online application area. Typical CF algorithms explores memory, model and hybrid based are introduced. Features like, listing out the traditional CF algorithms with their benefits and their limitations are described. An initial or common framework for CF recommender system is proposed in order to demonstrate that how CF recommendation takes place for several application areas with a sample problem. Hope so that the proposed initial framework will be useful to a novice researcher who is going to start their research in this field. In future data sparsity problem is going to analyzed and a new framework will be generated and validated depends on the user behaviors and user ratings respectively.

References

[1]. Young Sung so, song chul moon. "Frequent pattern to promote sale for selling associated items for recommender in E-Commerce". Indian journal of science and technology. oct 2016, vol9 (38), DOI: 10.17485 /ijst / 2016 / v9i38 / 102552.

[2]. Melville.P, Sindhwani.V 2010. "Recommender Systems" Encyclopedia of Machine Learning. Springer Science & Business Media.

[3]. J. Delgado, N. Ishii. "Memory-based weighted majority prediction. SIGIR Workshop Recomm. Syst. Citeseer, 1999.

[4]. B. Sarwar, G. Karypis, J. Konstan, J. Riedl, Item-based collaborative filtering recommendation algorithms. Proceedings of the 10th International Conference on World Wide Web, ACM. 2001.

[5]. Heng-Ru Zhang, Fan Min. Three-way recommender systems based on random forests. Elsevier, Knowledge-Based Systems. 2016, 91, pp. 275–286.

[6]. X. He, F. Min, W. Zhu. Comparison of discretization approaches for granular association rule mining. Can. J. Electr. Comput. Eng. (2014) , 37 (3) , 157–167.

[7]. B. Mobasher, R. Burke, J.J. Sandvig. Model-based collaborative filtering as a defense against profile injection attacks, in: AAAI, vol. 6, 2006

[8]. H.-R. Zhang, F. Min, Xu. He. Aggregated recommendation through random forests. Sci. World J. 2014, 1–11.

[9]. H.-R. Zhang, F. Min, S.-S. Wang, A random forest approach to model-based recommendation, J. Inf. Comput. Sci. (2014).

[10]. Le Hoang Son Dealing with the new user cold-start problem in recommender systems: A comparative review. jun 2016, Vol. 58, pp. 87–104,

[11]. Joao vinagre, Alipio , Mario Jorge. Forgetting Mechanisms for Scalable collaborative filtering. Springer. 2012, vol. 18, pp. 271-282.

[12]. Alfredo cutolo, Giuseppe D Aniello, Francesco orciuoli, francesa pettinati, Giuseppe sansonetti , catello vitagliano. An ontology – based recommender system in E-Commerce. 2nd international workshop on Recommender systems meet Big data & semantic technologies. May 2015.

[13]. Wu zen, zhang Xiaonan , Du yannan. A collaborative filtering recommender system integrated with interest drift based on forgetting function. International journal of u – and e-service, science and technology. 2015, vol.8, no.4, pp.247-264.

[14]. Zhe yang, Bing wu, kan zheng, xianbin wang, Lei Lei. A survey of collaborative filtering – based recommender system for mobile internet applications. IEEE translation and content mining. , May 2016,vol.4, pp.3273-3287.

[15]. Chen li , Zhengtao jiang. A hybrid news recommendation algorithm based on users browsing path, ICIS. IEEE Access. June 2016, pp.26-29.

[16]. Fuli zhang. A personalized time – sequence – based book recommendation algorithm for digital libraries. IEEE Translation and content mining. May 2016, vol.4, pp. 2714-2720.

[17]. Bushra Alhijawi , Yousef kilani. Using genetic algorithms for measuring the similarity values between users in collaborative filtering recommender system. ICIS. IEEE/ACIS 15th international conference. Aug 2016, Pp.26-29.

[18]. Zhipeng zhang, yasuo kudo , tetsuya murai. Neighbor selection for user – based collaborative filtering using covering – based rough sets. IUKM. Springer. Nov 2016, pp.1-16.