International Journal of Engineering Studies.
ISSN 0975-6469 Volume 8, Number 2 (2016), pp. 159-166

http://www.ripublication.com

# Semantic-based Web Mining Approach for Solving First Rate and Sparsity Problems in Recommendation Systems

**R. Shanthi [a], Dr. S.P. Rajagopalan [b]**

*[a]Research Scholar, Sathyabama University,*
*Rajiv Gandhi Salai, Chennai-600 119, Tamil Nadu, India.*
*E-Mail: sha_raju2003@yahoo.co.in,*

*[b]Professor, G.K.M. College of Engineering and Technology,*
*G.K.M. Nagar, Chennai-600 063, Tamil Nadu, India.*
*E-mail: sasirekaraj@yahoo.co.in*

**Abstract**

Web recommender systems are used in different domain applications to predict customer preferences. They also used to assist the web users to satisfy their search in different products and services. Nowadays, a web system with effective and reliable recommendation techniques has become the target of the research in recent years. Number of methods has been proposed to the users for effective and efficient recommendations ranging from traditional approaches to the recent sophisticated web mining techniques; still some drawbacks are present in the current recommender systems. A frame work have been proposed to improve the scalability and sparsity of the recommendation techniques, starting from collaborative filtering approach to the first rater and cold start problems. The frame work is addressed to any on-line product recommendation system and can be used or extended to other products/domains. It uses different predictive model techniques for recommendations based on different situations. The input data for these techniques are the product attributes and the user attributes integrated with various data mining algorithms structured according to particular domain ontology.
In this paper, a recommendation framework is proposed to overcome the major drawbacks of the current recommender systems in buying products.

**Keywords:** Semantic recommendation system, first-rate problem, sparsity problem.

**Introduction**

Internet and on-line activities have been increased during recent years. This has led the companies to know more about their consumer's behavior, so that they can satisfy their customers based upon their needs. A recommendation frame-work is proposed in order to improve the customer satisfaction. The framework is based upon the semantic web mining approach which is already applied to movie recommendation system and it is now extended to any on-line product recommender systems. In this paper recommender system is explained which is followed by associative classification and at last a frame work have been presented

**Related work:**

Recommender system methods for the customers are classified into 2 types as, Collaborative filtering (CF)approach and Content based approach(CBA).The first category is based on nearest neighbor algorithms which is used to predict product preferences for a user based on the opinions of the other users. The opinions are got by rating scores of the users. The content based approach is based on the contents /pages visited by a particular user. The drawback in this approach is the mechanisms to handle web objects such as motion pictures, images, music, large number of words etc. Two approaches are used for collaborative filtering algorithms

- Memory based/user based collaborative filtering
- Model based/item based collaborative filtering

Memory based algorithms otherwise known as nearest neighbor methods treats all users with statistical techniques to find users with similar preferences. The recommendation is predicted for an active user based on the neighborhood features. Pearson correlation coefficient technique is the popular technique used to find the prediction in which the prediction is computed by taking the weighted average of the product ratings of the nearest neighbors. The merit of this algorithm is quick incorporation of the most recent information and the drawback is search for the neighbors in large databases.

**Limitations in Memory Based Approach**

The memory based approach in collaborative filtering algorithm has the following limitations:

**Scalability:** is the problem which occurs where the number of ratings required is more than the number of ratings obtained.

**Sparsity:** is the problem which occurs when the search of neighbors in memory based collaborative algorithm. These problems are caused because there is a need to process large amount of information, when the number of products grows in size. To overcome this problem model based collaborative filtering is used. This method is based upon the user ratings to predict user preferences.

**Model Based Approach**

In model based method the time required for recommendation is less which is the major advantage of this approach. This is because the model is built offline before the user goes into the system and is applied on-line to facilitate the active user to go for recommendations. In model based method the time spent in building the model does not affect the response time of the user. In model based method, most recent data cannot be updated which is the advantage in case of memory based method.

The other problem occur in case of buying an on-line product is

- First rater/early start problem.: This problem arises when no recommendations are offered about an item that has just incorporated into the system which has only few evaluations from the users

**Associative classifications**

Sparsity problem in recommendation systems can be reduced by using associative classification algorithms such as class association rules(CARS) which is based on the class attributes classification based association (CBA). This algorithm is used which has 2 parts

- A rule generator based on association and
- Classifier builder

CMAR classification based on multiple class association is also used to solve sparsity problem.

FP growth algorithm is used instead of a-priori algorithm. Associative classification methods are not widely used in recommender systems even though they are well suited for sparsity problems.

**Semantic web mining**

Semantic web mining is used in web mining system to improve the recommendations for the customers. Humans only can understand the data which has no structure where the machine cannot. But only a machine can process large volume of data. Web mining is the knowledge extracted from the huge amount of web data. Semantic web enriches the web data with machine understandable information. Semantic web mining improves the results obtained by using web mining techniques.

In semantic web mining, ontology is used to describe the conceptualization of the data to be mined which is used to gain knowledge.

**Ontology**

In web mining data collection is the preprocessing stage of collecting data from heterogeneous databases and is to be integrated for knowledge discovery. Ontology uses a concept called taxonomic abstraction to gain meaningful results. Taxonomies are used to generate patterns at the abstract level. i.e., regularities are found between categories of products instead of specific products and these regularities/patterns are used in recommender system for recommending the new products.

When a new product is included in the catalogue no recommendation can be done when classical collaborative filtering algorithms are used. But, semantic web mining

which uses taxonomies can add new products and can be categorized and also recommendations can be made which is the way of addressing the first rate problem.
Application of traditional data mining methods can be extended to web systems in order to obtain benefits as obtained by using association and clustering techniques. Complementary products gives better results compared to products of similar types. Semantic information can be used in any domain to improve the quality of the recommendations with an ontological structure.

**Proposed recommendation system**

The proposed framework is used to overcome the major drawbacks of recommender systems such as Scalability, Sparsity and First rate problem.
All the drawbacks were dealt separately and no framework was proposed to handle all the problems jointly. Data mining algorithms generate predictive models for recommendations. Generation and the applications of the models are done by two separate process in which each process has 2 models.

*Off-line process*: this process is done by using induction of the predictive models and is done before the users are using the system(offline). These models are updated periodically inorder to add the most recent information regarding the new users and new products as well as the ratings given by the users to the products.

*On-line process:* this process is carried out when users request a recommendation (on-line). This process varies for new and old users because preferences about the new user are known but preferences about the new users are not known because they have not rated any product again off-line process constitute in building 2 different models for recommendations

1. Low level model : recommendations are made in ordinary situations related to particular products and users in preference ratings . no semantic annotations are used
2. High level model: this model is related with the types of products and types of users for ratings classification of products and users should be done previously based upon ontology for particular application area.

Recommendations are done based on the characteristics of both the users and the products but not on their evaluations. High level models are applicable to solve first-rater problem which arise in recommendations. Both low level and high level models are used during on-line process at recommender time. Recommendations are provided only to the registered users in on-line.
On-line recommendations are applicable for both old and new products and for the old and new users. Most recent products cannot be added in the case of low-level modelbecause it suffers in first-rater problems but on the high level model it is possible for the recent products to make classifications according to their characteristics. Ontology is used to classify the new users where the new users are to be registered to get the semantic information about them.

**Advantages in building off-line model.**
It avoids scalability since the time spent on induction for the models does not affect the response time of the user because user always expect fast feedback. Scalability does not take place in the proposed framework, because data is not processed at recommendation time.
Sparsity problem is reduced by using associative classification methods. Associative model-based methods reduces the sparsity complexity. An associative classification algorithm is better compared to other classification algorithms for reducing sparsity. The data taken into account for the application domain is

**User:** user_id, gender, age, occupation

**Product**: product_id, product name, product category, product company, product brand, purchase date, product size, product measure purchase quantity, purchase amount, offered, offer/price value.

**Rating:** rating_id, user_id, product_id, score, rating_value.
Associative classification is the better behavior of the method to solve sparsity problem. More recommendations can be made with a lesser number of ratings. The classifications algorithms used for the study are of two types

1. Associative classification algorithm
2. Non-associative classification algorithm

**Associative classification algorithm**
The following associative classification algorithms are used to compare the efficiency among them in product recommender systems.

- CBA- classification based on association
- CMAR-classification based on multiple class association rules
- FOIL- first order inductive learner.
- CPAR-classification based on predictive association rules

**Non- Associative classification algorithm**

- Decision tree
- Bayes net
- Nearest neighbor
- Random tree

**Table 1:** Precision value for Non-associative classification algorithms

| Associative algorithms | Precision (%) |
|---|---|
| CMAR | 91 |
| CBA | 85 |
| FOIL | 84 |
| CPAR | 49 |

For the above given table 1 the chart representation indicates, of all the given ASSOCIATION algorithms FOIL is the only method to reduce the computational cost and also increased the accuracy value. The best precision is obtained by classification based on multiple class association rules.

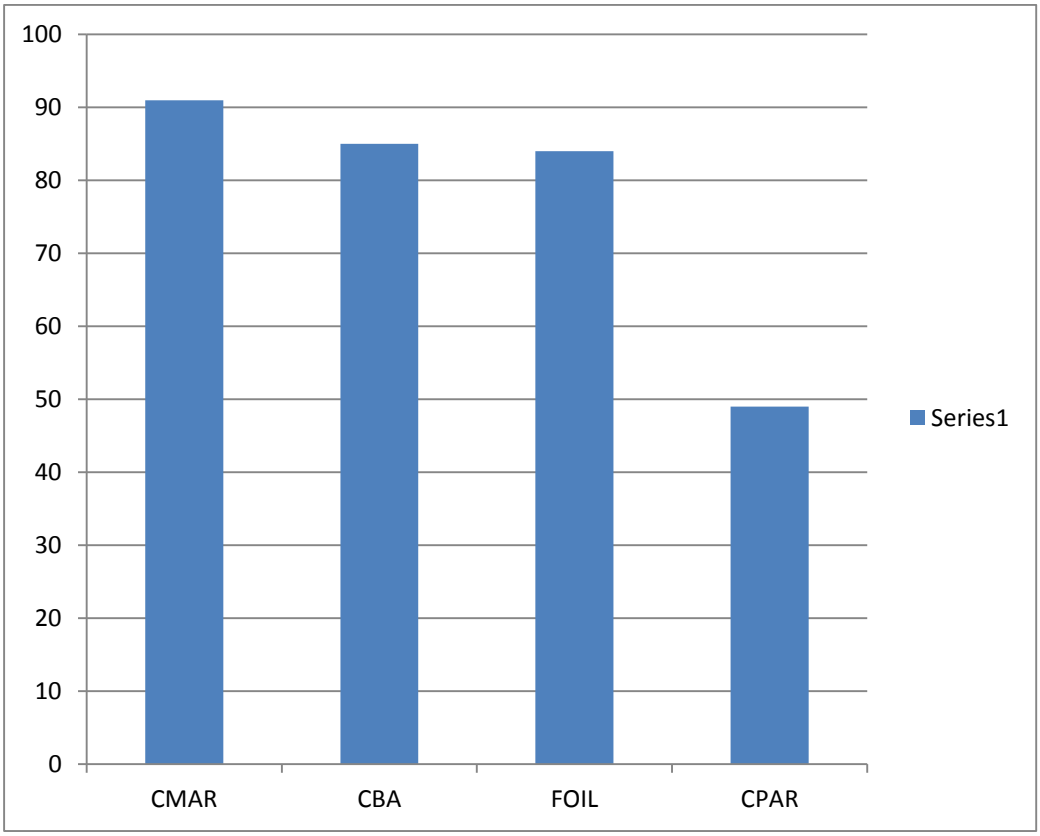
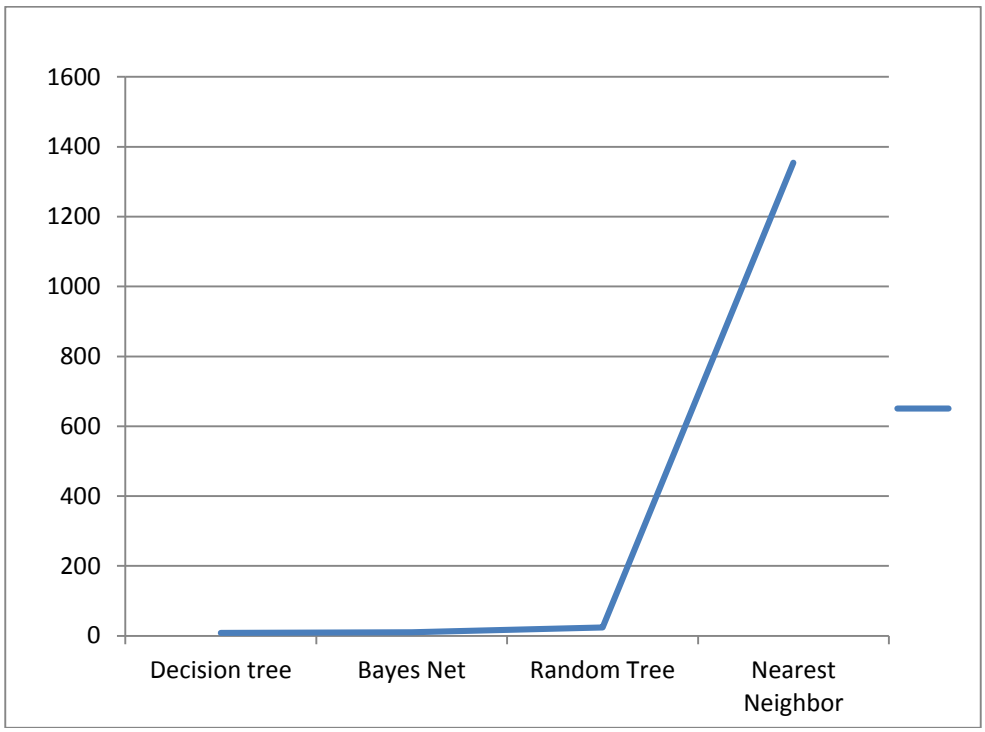



**Figure 1:** Precision for different associative classification algorithms

Execution time for Non-associative classification algorithms

| Non-Associative algorithms | Execution Time in seconds (Multi classifiers) |
|---|---|
| Decision tree | 8 |
| Bayes Net | 10 |
| Random Tree | 24 |
| Nearest Neighbor | 1354 |




**Execution Time for different Machine learning algorithms**

Compared to associative classification and non-associative algorithms, association classification algorithms reduced the errors occurred in non-associative classification algorithms.

## Conclusions

The recommender system is proposed in choosing an on-line product to a customer. This frame work reduces the complexity of sparsity and first rater problem which may occur in choosing the products. The off-line model avoids scalability problem and the proposed technique which uses associative classification methods provides a way to deal with sparsity for choosing products for the on-line customers.

## References

[1] M.Claypool, A.Gokhale, T.Miranda, P.Murnikov, combining content-based and collaborative filters in an online newspaper, A Proceedings of the ACM,1999.

[2] Tong, Wang and Pi-lian, He, Web Log Mining by an Improved AprioriAll Algorithm World Academy of Science, Engineering and Technology, Vol 4 2005 pp 97-100.

[2] Dunham., Margaret H., Data Mining Introductory and Advanced Topics.

Beijing: Tsinghua University Press, 2003, p195-220.

[3] Han Jiawei and Kamber Micheline Data Mining Concepts and Techniques [M].Beijing: China Machine Press, 2001, p290-297.

[4] Bain Tony SQL Server 2000 Data Warehouse and Analysis Services. Beijing: China Electric Power Press, 2003, p443-470.

[5] Bin, Lin jie, de, Liu ming and xiang, Chen Data mining and OLAP Theory & Practice [M]. Beijing: Tsinghua University Press, 2003, p194- 244.

[6] Tong, Wang and Pi-lian, He, Web Log Mining by an Improved AprioriAll Algorithm World Academy of Science, Engineering and Technology, Vol 4 2005 pp 97-100.

[7] Wen-Hai Gao Research On Client Behaviour Pattern Recognition System Based On Web Log Mining, Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, Qingdao, 11-14 July 2010, DO1, 978-1-4244-6527-9/10/$26.00 ©2010 IEEE, pp 466-470

[8] Levi Albert, and Koç Çetin Kaya CONSEPP: Convenient and Secure Electronic Payment Protocol Based on X9.59 Proceedings, the 17th Annual Computer Security Applications Conference, pages 286-295, New Orleans, Louisiana, IEEE Computer Society Press, Los Alamitos, California, December 10-14, 2001

[9] Guo, Di, Collector Engine System: A Web Mining Tool for E-Commerce, Proceedings of the First International Conference on Innovative Computing, Information and Control (ICICIC'06) DOI computer society 0-7695-2616-0/06 2006 IEEE Yuewen, LI, Research on E-Commerce Secure Technology DOI 978-1-4244-3709- 2/10 2010 IEEE

[10] Mei Li and Cheng Feng, Overview of WEB Mining Technology and Its Application in E-commerce, 2010 2nd International Conference on Computer Engineering and Technology ,Volume 7.DOI 978-1-4244- 6349-7/10 .2010 IEEE pp V7-277-V7-280

[11] Yue, Chuan, Xie, Mengjun, and Wang, Haining, Automatic Cookie Usage Setting with Cookie Picker, 37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN'07) DOI 0-7695-2855-4/07 2007 IEEE