

A Survey on Recommendation Techniques in E-Commerce

Namitha Ann Regi
Post-Graduate Student
Department of Computer Science and
Engineering
Karunya University, India

P. Rebecca Sandra
Assistant Professor
Department of Computer Science and
Engineering
Karunya University, India

Abstract

Recommendation technique is a method for suggesting products to the customers in large e-commerce where there are plenty of products available. The items recommended should have a high degree to be liked by customers. Recommendation can be provided to customers by predicting the rating he/she can give to a particular item. Product can be recommended to a customer if the estimated rating is high. Collaborative filtering, content-based filtering, association rule mining are some of the approaches used for recommendation. Neighbourhood method and latent factor models are the two approaches of collaborative filtering. Item oriented and users oriented are the two types of neighbourhood model. This paper does a comparative study for some of the recommendation techniques used in e-commerce.

1. Introduction

Extracting data from large and huge amount of data is called data mining. Knowledge is discovered from data in which essential data is taken out from the raw material called knowledge discovery from data. Data mining is the process of analyzing the data from different point of view and classifies it into different and separate groups. In the field of e-commerce, huge amount of items and products are available and it makes users difficult to choose the best item that having the features according to their interest. Thus selecting item is a time consuming process. The problem can be solved by recommendation techniques. The technique is used to recommend items or products to users according to their own interest so that it is easy for users to select items and save their time. This technique calculates the possible rating that can be given to a particular item by particular user. Items can be recommended to a customer according to the rating estimated.

Varieties of techniques are available for recommendation having their own advantages and disadvantages. Root mean square error is method used to evaluate the error in the estimated rating. Difference between the estimated rating and real rating are calculated and find out how accurate the technique is. Collaborative filtering, content-based filtering, association rule mining are some of the approaches used for recommendation. Collaborative filtering uses the interest and opinion of other users and information about the past behaviour of the user. Neighbourhood method and latent factor models are the two approaches of collaborative filtering. Item oriented and users oriented are the two types of neighbourhood model. Matrix factorization is a technique used in latent factor models in which latent factors are learned from the pattern of rating. This paper does a comparative study for some of the recommendation techniques used in e-commerce.

This paper is organized as follows. Section 2 includes a discussion on key concepts presented in this paper. Section 3 is a comparative study of recommendation techniques in e-commerce. Section 4 gives the conclusion of this paper.

2. Key concepts

2.1. Data mining

Extraction of essential data from bulk and huge amount of data is called data mining. Knowledge is discovered from data in which essential data is taken out from the raw material called knowledge discovery from data. Data mining is the process of analyzing the data from different point of view and classifies it into different and separate groups.

2.2. Recommendation technique

Huge amount of items and products are available and it makes users difficult to choose the best item that having the features according to their interest in the

field of e-commerce. The problem can be solved by recommendation techniques. The technique is used to recommend items or products to users according to their own interest so that it is easy for users to select items and save their time. This technique calculates the possible rating that can be given to a particular item by particular user. Items can be recommended to a customer according to the rating estimated.

3. Analysis of different recommendation techniques

This section contains a study on some of the recommendation approaches in e-commerce.

3.1. Item-Based Collaborative Filtering Algorithm.

Collaborative filtering approach [1] done predictions based on interest and opinion of other users. The amount of work increases as the members who take part increases thus more effort has to be taken. Item based recommendations are introduced to produce good quality recommendations for wide range of users. Item based technique is done based on the relationship between items and provides recommendations. The approach involves two steps to estimate the rating a user 'u' is going to give to an item 'i'. The first step is to find out the items the user 'u' has rated already and calculate the similarity between those items and selected item 'i'. The second step is to select the most similar items and use an algorithm to acquire prediction using the ratings given to the similar items.

Cosine-based similarity, correlation based similarity and adjusted cosine similarities are the three techniques to calculate the similarity between two items. Cosine-based similarity is calculated using the cosine of the angle between vectors of the items. This can be done by dot-product method. Correlation-based Similarity separates the users who have rated for both items and calculate the similarity between them. Adjusted cosine similarity is similar to cosine based approach in which it calculates the difference between the user average and co-rated pair. Similarity is computed using the columns of the matrix. Rows of the matrix are considered for user-based collaborative filtering.

Weighted sum and regression are two approaches to acquire prediction. Weighted sum is a technique in which it selects the most similar items and gets the summation of the ratings given to these items by the target user. Regression technique gets the approximation of the ratings given to the most similar

items in which it does not take the exact values of rating.

The main advantage of Item-based algorithm is that it provides finer quality and effective performance than user-based algorithms but it is difficult to estimate rating for a new user in classic collaborative filtering approach because the user has not given any rating on the items.

3.2. Recommendation on sparse Data

Accurate prediction of rating [2] is prevented by two problems. They are sparsity and dynamic nature. Dynamic nature of users is changing the user preferences in multiple phase of interest. The variable does not contain the actual value is called sparsity. Otherwise rating values are missed in the user-item matrix. These issues reduce the coverage and accuracy. Dynamic personalized recommendation algorithm provides accurate prediction of rating when sparsity and dynamic nature exist. The algorithm involves three steps. They are relation mining of rating data, dynamic feature extraction and adaptive weighting algorithm.

Relation mining of rating data involves mining of data from user profiles and item profiles and find out the relation between indistinguishable content in one or more attributes. User profile contains information such as user id, name age, date of birth, occupation, and region and so on. The item profile consists of information such as item id, genre. Select the data set whose rating is semi-co-rate related with candidate rating. Classification and clustering are the two techniques that are performed in each attribute of user profile and item profile and it find out the neighbouring ratings.

Dynamic feature extraction finds out the drifting of signals as fast as possible and to be updated frequently. The approach divides the rating received from relation mining of rating data into several disjoint secondary subset. Feature extraction is done by performing time series analysis algorithm on secondary subset. Time series analysis is used to outline user's preferences and item's reputation. It calculates the difference between the timestamp of candidate rating and each rating in the secondary subset.

Adaptive weighting algorithm uses time and density factors to calculate the estimated rating and calculate the difference between estimated rating and real rating to get the error in computation. Prediction is accurate when the error is least. Root mean square error (RMSE) is used to evaluate the algorithm

The main advantage of the algorithm is that it is effective when the data is changing frequently and dynamic and produce better performance than previous algorithms.

3.3. Integrating Content Based and Collaborative Filtering

Recommender system [3] provides a way to extract and suggest information to the users from large information spaces of e-commerce. Users can meet their requirements through the recommender systems. Content based and collaborative filtering are two techniques used in the field of recommender systems. Both the techniques have their own advantages and disadvantages. Content based approach uses semantic content to select the information and collaborative filtering selects information based on the interest and opinion of other users. A hybrid approach that combines both the content based and collaborative filtering approach provides better performance and utilizes the strength of two techniques such as content based and collaborative filtering.

The hybrid approach consists of four steps in which first step is to create profile for all the users. User profile should consist of information such as several attributes and weights on the interested items. User profiles are created to get the semantic content which is used for similarity calculation. Manual weighted method and auto weighted methods are the two approaches to create the user profiles. The users provide information such as interested items and how much they like the item.

The second step is to create a group rating matrix. The user profiles can be grouped using clustering algorithm such as K-means algorithm and fuzzy k-means algorithm. Certain number of clusters is generated for user profiles. The object should be a member of one of the clusters. Fuzzy k-means is used to class together the items and an object receives a fuzzy membership during its iteration.

The third step is to compute the similarity. Adjusted-cosine algorithm and Pearson correlation algorithm are the two techniques used for similarity computation. Adjusted-cosine algorithm calculates the sub user-user similarity for group rating matrix and Pearson-correlation algorithm calculates sub user-user similarity for user rating matrix. User-user similarity is calculated using the above two.

The final step is performing the prediction collaboratively for each item. Generate top-N rule and

according to the similarity select the nearest N neighbours.

The main advantage of this algorithm is that more information is extracted. It utilizes the advantages of content based and collaborative filtering. Filtering is better.

3.4. Association Rule mining, Decision Tree Induction

The customer [4] gets difficulty in choosing products from e-commerce sites due to large amount of data and product availability. Recommender systems provide a way to overcome these problems. The techniques used to overcome these problems are web usage mining, product taxonomy, and association rule mining and decision tree induction. Participants of the group get recommendations based on the past actions of the group members. The approach consist of four steps such as Finding out customers those who are active, Discovering affinities between the products, Determining first choice of the customer and making recommendations.

The active customers are determined using decision tree induction technique. It is not needed to provide recommendation for user who does not buy recommended products. The customers are active selected in this step. Model set and score set are used for decision tree technique and it they are produced from customer data. Model candidate set consist of customers who are present in the model set and Score candidate set consist of customers who are present in score set. Time frame is used with model set to get effective model. It consist of past, current and future data.

Affinity between products is exposed in the next step. Association rule mining is the procedure used to determine product affinity. Association rules are generated for basket placement transaction set and purchase transaction set.

The third step is to discover preference for active customers. Previous purchase details are tracked to find out the behaviour of their shopping. Click-through and basket placement are used to construct customer preference model.

Predicting recommendation is the fourth step in which product appeal model and customer preference model described in the above are used to make recommendations.

3.5. Model for the Purchase Probability of Anonymous Customers

It is difficult to provide [5] recommendations when customers do not use personal login and navigate web pages. The methods used to overcome these problems are extracting purchase patterns and predicting purchase probability. Web marketers get marketing suggestions from purchase patterns and Purchase probability is predicted for anonymous customer using predicting purchase probability technique.

Phase 1 consists of five steps. They are web log data, Data pre-processing, mining association rules, selecting purchasing patterns and marketing implications. Phase 2 also consists of five steps such as data set, modelling data set, training prediction models, predicting purchase probability.

Web log data gathers data of users of the web. Number of hits and associated information are elements of web log data. Associated information includes number of transferred bytes, URL, cookie data, IP address, customer agent, timestamp.

Unnecessary information is removed from web log data through data pre-processing. IP address, URL, time and session are the four elements that are selected to remove unnecessary information. Navigation of a customer can be determined using the attributes such as IP address, time and session. Client transactions are discovered from URL. URL is transformed to keyword to take out necessary information.

Mining association rule is the third step in which it uses a priori algorithm. A priori algorithm consists of two steps. Downward closure property of support is used to produce all transaction set. Rules are produced from the set of all frequent item sets.

Patterns are selected from discovered association rules. Redundant rules and trivial patterns should be removed from association rules and select meaningful patterns having last path as order.

Purchasing pattern discovered in the phase 1 is used to produce attributes of modified data set and transforms web log data to data set for data mining algorithm. Association rule mining and purchasing patterns converts original data set into modified data set. Original data set consists of IP address, session, access time, URL and modified data set consists of IP address, Target address, considerable web pages and series of web pages.

Decision tree, artificial neural network and logistic regression are three prediction models to predict future values. A cross approach of these three models is used to train modified data set. Input data is same for three prediction models.

The combination of above mentioned algorithms such as decision tree, artificial neural network and logistic regression are used to predict purchase probability. It provides best accuracy. Accuracy is computed in the first step, purchase probability is predicted from independent classifier in the second and third step. Finally to minimize misclassification, decide upon θ parameter.

The main advantage of the algorithm is that purchase probability is predicted for anonymous customers and provides product recommendation and good customer inducement.

Table 1. Comparative Study of Different Recommendation Techniques

Technique	Method	Merits
Item-Based Collaborative Filtering Algorithm	Item-item similarity and recommendation for users	(i) Better performance than user-based. (ii) Quality of prediction is good
Dynamic Personalized Recommendation On Sparse Data	Relation mining of rating data and Feature is extracted dynamically using time series analysis	The problem of sparsity and dynamic nature are solved.
Integrating content-based & collaborative filters	Combines content based & collaborative filters	(i) More information is extracted. Exploit the advantages of CF & CBF. (ii) Better filtering.
Association rule mining, Decision Tree Induction	incorporate product affinities and customer preferences	Provides good quality recommendation based on user behaviour
Model for the purchase probability of anonymous customers	Extracting purchase pattern and predicting purchase probability	(i) Small amount of data is needed. (ii) Customers without membership also can apply. (iii) Processing time is very less to calculate probability

4. Conclusion

Some of the best recommendation algorithms in e-commerce have been discussed in this paper. All these algorithms focus on increasing their performance and accuracy. Main advantages and disadvantages of these algorithms are also described in the paper. This paper will aid to take decisions on which algorithm to be used for recommendation in different situations or scenarios.

5. References

- [1] B. M. Sarwar, G. Karypis, J. A. Konstan, J. Riedl, "Item-based collaborative filtering recommendation algorithms", Springer Science, 2006.
- [2] X. Tang, Jie Zhou "Dynamic Personalized Recommendation on sparse data", IEEE Transaction on knowledge and data engineering, 2012.
- [3] B. M. Kim, Q. Li, C. S. Park, S. G. Kim, J. Y. Kim, "A new approach for combining content-based and collaborative filters", 2006.
- [4] Jae Kyeong Kima, Yoon Ho Chob, Woo Ju Kimc, Je Ran Kimc, Ji Hae Suha, "A personalized recommendation procedure for Internet shopping Support", Electronic Commerce Research and Application, 2002.
- [5] Euiho Suh, S Lim, H. Hwang, S.kim, "A prediction model for the purchase probability of anonymous customers to support real time web marketing", Elsevier, 2004

IJERT