International Journal of Modern Electronics and Communication Engineering (IJMECE) ISSN: 2321-2152
Volume No.-7, Issue No.-3, May, 2019

A Collaborative Recommendation System for e-shopping using URRP with LDA

Lavanya.S.C
Department of Computer Science and Engineering
Dr.Mahalingam College of Engineering and Technology
Coimbatore, India
E-mail: lavanyack5@gmail.com

Gowrisankari.T
Department of Computer Science and Engineering
Dr.Mahalingam College of Engineering and Technology
Coimbatore, India
E-mail: shankarii.g@gmail.com

Suganya.B
Department of Computer Science and Engineering
Dr.Mahalingam College of Engineering and Technology
Coimbatore, India
E-mail: suganyagovindharaaj@gmail.com

Abstract: Recommendation System has been extremely common in recent years and also has changed the trend how people search for products. Reviews accompanied by ratings provide better information to user for buying any product. Most recommendation systems works only on ratings accompanied by positive reviews and generate scores for granted and discard the useful information in accompanying reviews. Collaborative Filtering (CF) technique with Latent Dirichlet Algorithm (LDA) based topic model considers user attitude that enables to identify useful information about user profile embedded in both ratings and reviews (both positive and negative). User Ratings and Reviews Profiling (URRP) identifies good products by constructing profiles based on the ratings and reviews given by the user.

Keywords: Recommendation systems, Latent Dirichlet Algorithm, User Ratings and Review Profiling, Collaborative Filtering.

I. INTRODUCTION

In the past years, the number of personalization applications has strongly increased, especially in the field of electronic commerce where personalization becomes an important success factor. The term personalization [1] means the filtering of information for each particular person in order to provide customers a customized or personalized interaction with a company's products, services, web site and employees. The starting point for collaborative filtering is an m-by-n-matrix (called rating matrix) with m referring to customers (rows) and n referring to products (columns). By using different techniques, the similarities between the products (item-based technique) or between the users (user-based technique) are calculated.

The information used to fill the rating matrix can either be gained explicitly or implicitly. Explicit information is entered by the customer directly, whereas implicit information is retrieved from the user's interaction with the shop. Explicit information includes product ratings given by the customer. Implicit information includes the orders and the clickstream analysis. The collaborative filtering method is a very efficient and convenient way of achieving personalization, as there is no need to introduce semantic information about the products or to manually link products and users together. The customer's interaction with the shop is the only required information.

Recommendation Systems are used to give recommendations to the user according to the ratings and reviews given to the product by other users. Now-a-days, Recommendation System is much useful for the user to know which product has better specifications when compared to similar products. It is also helpful for a new user to decide which product is best, though a new user has no idea about the particular product.

Collaborative filtering uses the past customer behavior for recommending the products. It also makes recommendations based on the similarities of the customers. This approach is used to predict whether the user is interested in that product or not. CF technique [2] recommends products based on other people who have similar tastes with target users. Two limitations with CF technique are cold start problem and sparsity problem. Limitations can be resolved using new personalized recommendation model, i.e. Topic Model based Collaborative Filtering (TMCF) utilizing users' reviews and ratings. Extended LDA model [2] to generate topic allocations for each review is used and then each user's preference is obtained.

Review text content is an important source of information for obtaining personalized information regarding users. The authors of [8] considered the personalized information of micro-blog users, proposed a personalized micro-blog sentiment classification method, and achieved better sentiment classification performance. The user-item rating matrix is another data source for obtaining personalized information about users. From the perspective of the recommendation system, based on the historical rating in the user-item rating matrix, the personalized information of the users can be mined through the collaborative filtering algorithm

II. RELATED WORKS

Content based filtering recommends product or item to a user based on the users past behaviour by analyzing their profiles, ratings and reviews. It also recommends a product by matching up the similarities between the products based on user preferences. In order to find whether the user likes or dislikes a new product, user's implicit tastes need to be analyzed. For example, some users like action movies and some users like romantic movies. So both tastes of the user as well as the characteristics of the product need to be predicted. For this purpose both latent ratings and latent review text of the user have to be considered.

The collaborative filtering approach [1] can be implemented using user-based or item-based methods. Both take as input the rating matrix with the customers in the row dimension and the products in the column dimension. This two-dimensional matrix represents the relationships between users and products either based on product ratings, purchased products or clickstream data. If product ratings are considered, each element at the intersection of a product and a customer will contain a value between -1 and +1 representing the judgment of the customer for the product where -1 denotes a strong dislike and +1 a strong affection as shown in table 1.

Table1. Similarities between customers

	Raj	Sam	Peter
Raj	-	-1	+1
Sam	-1	-	-0.7
Peter	+1	-0.7	-

The collaborative filtering method is a very efficient and convenient way of achieving personalization as there is no need to introduce semantic information about the products or to manually link products and users together.

In order to recommend products to users there is a need to predict how a user will respond to a new product. To do so it is necessary to uncover the implicit tastes of each user as well as the properties of each product. User feedback is required to discover these latent product and user dimensions. Such feedback often comes in the form of a numeric rating accompanied by review text. However, traditional methods often discard review text, which makes user and product latent dimensions difficult to interpret, since they ignore the very text that justifies a user's rating. The author of [3] aims to combine latent rating dimensions (such as those of latent-factor recommender systems) with latent review topics (such as those learned by topic models like LDA). Highly interpretable textual labels for latent rating dimensions, is used to 'justify' ratings with text and it also more accurately predicts product ratings by harnessing the information present in review text; this is especially true for new products and users, who may have too few ratings to model their latent factors, yet may still provide substantial information from the text of even a single review. The discovered topics can be used to facilitate other tasks such as automated genre discovery, and to identify useful and representative reviews.

The User Rating Profile model can be developed, providing a Gibbs Sampling [4] derivation for parameter estimation. Gibbs Sampling provides a simple and flexible learning procedure which can be extended to include external evidence, in the form of soft constraints. With a-priori information about user-neighbors, an effective regularization technique that drives the first sampling iterations pushing the model towards a state which better represents the user-neighborhoods specified in the input. Collaborative filtering and content-based filtering is combined to learn user rating and review preferences more accurately.

The existing recommendation system produce recommendations based on the ratings and reviews (only positive) given by the user. Because of considering only positive reviews, accuracy is less and it need to be improved. Collaborative filtering uses the past customer behavior for recommending the products. It also makes recommendations based on the similarities of the customers. One of the most successful Collaborative Filtering approaches is Latent Factor Model [5] that recommends related products to users by combining the merits of latent factor model and probabilistic topic model such as latent Dirichlet allocation(LDA), aiming to learn latent user factors from observed reviews rating and latent items factors from reviews text. It provides an interpretable latent factor for users and items.

Collaborative filtering approach is used to predict whether the user is interested in any particular product or not. This is automatically done by considering the taste and age information of the user. This model predicts the behavior of any user by monitoring what the user has done as well as comparing the particular user's behavior with all other users. Latent factor model is implemented by applying matrix factorization techniques on the rating matrix. The matrix represents the relationship between user and product based on ratings. The accuracy of the recommendation system is improved by learning the latent factor models. In addition to Collaborative filtering, User rating profiling (URP) algorithm is used. It is an extension of Latent Dirichlet Allocation (LDA). It is a generative statistical model that allows sets of observations to be explained by unobserved groups that explain why some parts of the data are similar. The ratings are considered as topics and the attitudes for topics are decided. For each attitude, sampling probabilities and attitude distribution are computed. A sample attitude and rating value for each item is considered.

III. USER RATING AND REVIEW PROFILING

In addition to URP algorithm, the system also uses User Rating and Review Profiling (URRP) algorithm [6]. This algorithm is a combination of both User Rating Profiling (URP) and Latent Dirichlet Allocation (LDA). The basic idea is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words. In the working of LDA, each latent topic dimension is identified and a sample word distribution is made. Then for each review a sample topic distribution is made. For each word a sample topic assignment is made. The topics are assigned and from these topics the user's attitude can be determined. There will be 'M' users ,'N' number of items, 'K' number of latent topics or attitudes, W number of words, user topics or attitudes distribution, topics distribution over words, topic assignment for each word, topic assignment for each word excluding observations.

Recommender system search for inputs in the datasets. The inputs are text or string which is the item name. While searching the datasets, the attitudes (both reviews and ratings) are extracted from the input and assigned to its corresponding topic. After assigning the topic, the parameter is estimated for each and every attitude by performing Gibbs sampling. The Gibbs sampling is performed for updating new topics for the attitudes. The topics are randomly initialized for each and every word. The current attitudes and its related counts are excluded. For new attitude, a new sampling is made. A new value is assigned to the attitude. The new attitude's value is updated to its corresponding topic assignment. The topics which are unavailable are generated based on the new reviews from the user. The topics generated are updated and the corresponding values for the topics are also updated. The next step is to generate score for the items. The score is generated by assigning it to a new variable. The variable keeps hold of all the scores which are generated. The scores are specifically generated and stored in the database for each and every item. The score generation is applied on three datasets namely training, validation and test datasets. The training dataset is utilized for 80% of the process and the remaining operations are done on validation and test datasets. The scores were calculated by processing only the positive topics and attitudes in the reviews. The words which are marked good are compared with the words in the reviews. The words which match are denoted as good and the score is incremented. Similarly the total score is calculated for every individual item. The items are recommended based on these scores. The items with the highest scores are considered as good items and the results are recommended.

IV. METHODOLOGY

Recommender systems are already used by many e-shopping sites under various formats, interfaces and underlying technologies. Many e-shopping sites use recommendation system to enable buying, offer personalized customer service and build loyalty. The techniques used to support ranges from the manual definition of recommendations by a system administrator to the automatic generation of recommendation rules based on the user behavior. In order to recommend products to users the system must ultimately predict how a user will respond to a new product. To do so the system must uncover the implicit tastes of each user as well as the properties of each product.

Recommender systems [3] have transformed the way users discover and evaluate products on the web. Whenever users asses products, there is a need to model how they made their assessment, either to suggest new products they might enjoy, to summarize the important points in their reviews, or to identify other users who may share similar opinions. Such tasks have been studied across a variety of domains. The system [7] uses positive as well as negative reviews for generating the recommendation results. The system works by processing the shopping website dataset. In shopping dataset, a specific dataset is chosen and the preprocessing is done.

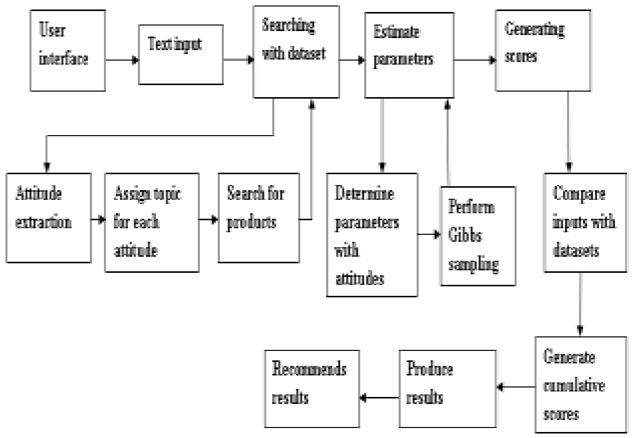

Figure1. Model of a recommender system

The system uses User Rating and Review Profiling algorithm. The topics are decided first and then a sample topic distribution is made. From topics, user's attitudes are extracted. By extracting the attitudes, a sample rating probability is done for each item. Then for each user a sample topic assignment is done and the corresponding attitude is stored. The topics from the reviews are matched with the attitudes with the help of Latent Dirichlet Association (LDA). The next step is to decide parameters for each and every attitude and its topic. This is done with Gibbs sampling algorithm. The topics with attitudes are estimated and then a distribution is done with the help of the sampling parameters. The similar attitude distribution are also sorted out and processed separately. For a new attitude, a new attribute is assigned. The words in the review are updated to a new topic. Then each and every attitudes, topics, words are updated. The score is then generated for the estimated parameters. The scores are generated by comparing the words in the reviews with a separate positive and negative dataset.

The positive and negative datasets have words which are listed under the topics good and bad. The words in the reviews are compared with dataset to check whether the review is positive or negative. The score for the positive review is incremented to the temporary score and for negative review the score is decremented with the temporary score. Finally a cumulative score comprising of both positive and negative scores is determined. The score for the product is calculated. The items are recommended in descending order based on the scores. Efficiency of the system is calculated using Mean Squared Error (MSE). The MSE has estimator and predictor. The estimator determines the average number of square of errors. The predictor maps arbitrary inputs to the sample probabilities of topics.

V. RESULTS AND DISCUSSIONS

URRP with LDA system exploits information embedded in ratings and reviews in order to make accurate recommendations [6]. It links traditional collaborative filtering with topic modeling seamlessly, by applying the same multinomial distribution to each user's latent rating factors and review topics, since user rating behaviors and review topics are essentially the same. URRP system utilizes the complete dataset, inclusive of both positive and negative reviews.

The results are compared using Mean Absolute Error (MAE). The MAE values of URRP System (Positive and Negative Reviews) system are mostly reduced when compared with the MAE values of URRP System (Positive Reviews). The lower the MAE, higher the efficiency of the system. The results obtained clearly shows the improvement of results of URRP with LDA system over the existing system. The efficiency percentage is increased to a considerable extent.

Table 2. Estimation of Efficiency of URRP systems

Dataset Name	No. of Reviews and Ratings	MAE for URRP System (Positive Reviews)	MAE for URRP System (Positive and Negative Reviews)	Improvement in Efficiency
Android Apps	251	0.988	0.999	-0.011
Anklet	325	1.465	1.232	0.233
Automotive	216	1.423	1.433	-0.01
Baby	136	1.425	1.369	0.056
Books	390	1.604	1.601	0.003
Mobile	336	1.557	1.534	0.023
Electronics	289	1.22	1.18	0.04
Music	196	1.012	0.983	0.029

Sports	223	2.019	1.992	0.027
Toys	178	1.413	1.423	-0.01

Table 3. Comparison of MAE of URRP systems

System	Average MAE
URRP (positive reviews)	1.412
URRP (positive and negative reviews)	1.375

The experimental results show the superiority of our system over the existing system. Also with review text information involved, latent user rating attitudes are interpretable and "cold-start" problem can be lessened.

VI. Conclusion

This system uses URRP with LDA that works by considering the user attitude. URRP links traditional collaborative filtering with topic model. By introducing the wealth of information in reviews, URRP can learn users' rating behavior more accurately even with the few ratings. This system identifies information about user profile embedded in ratings and positive, negative reviews. Considering negative reviews, reduces mean squared error value. The test results show that URRP with LDA by considering positive and negative reviews provides efficient recommendation than those of considering only positive reviews.

References

[1] Uwe Leimstoll, Henrik Stormer, "Collaborative Recommender Systems for Online Shops," 13th Americas Conference on Information Systems, 2007.

[2] Jingnan Xu, Xiaolin Zheng, Weifeng Ding, "Personalized Recommendation Based on Reviews and Ratings Alleviating the Sparsity Problem of Collaborative Filtering," IEEE Ninth International Conference on e-Business Engineering, 2012.

[3] Julian McAuley, Jure Leskovec, "Hidden Factors and Hidden Topics: Understanding Rating Dimensions with Review Text", Proceedings of the 7th ACM conference on Recommender systems, pp. 165-172, 2013.

[4] N.Barbieri, "Regularized Gibbs sampling for user profiling with soft constraints," International Conference on Advances in Social Networks Analysis and Mining, 2011.

[5] Jing Peng, Ying Zhai, Jing Qiu, "Learning latent factor from review text and rating for recommendation," 7th International Conference on Modelling, Identification and Control, 2015.

[6] Mingming Jiang, Dandan Song, Lejian Liao, Feida Zhu, "A Bayesian Recommender Model for User Rating and Review Profiling," Article in Tsinghua Science & Technology, 2015.

[7] Te Min Chang, Wen Feng Hsiao,"LDA-based Personalized Document Recommendation," Proceedings of The Pacific Asia Conference on Information Systems, 2013.

[8] Bingkun Wang, Bing Chen, Li Ma,Gaiyun Zhou, "User-Personalized Review Rating Prediction Method Based on Review Text Content and User-Item Rating Matrix," Information, 2019.

[9] Yifan Hu, Yehuda Koren, Chris Volinsky, "Collaborative Filtering for Implicit Feedback Datasets," Eighth IEEE International Conference on Data Mining, 2008.

[10] George Prassas, Katherine C. Pramataris, Olga Papaemmanouil, Georgios J. Doukidis,"A Recommender System for Online Shopping Based on Past Customer Behaviour," 14th Bled Electronic Commerce Conference, pp.766-782, 2001.

[11] HePing Li, Feng Zhang, ShuWu Zhang, "Multi-feature hierarchical topic models for human behavior recognition," Science China Information Sciences, Volume 57, Issue 9, pp. 1–15, 2014.

[12] Guang Ling, Michael R. Lyu, Irwin King, "Ratings meet reviews, a combined approach to recommend," Proceedings of the 8th ACM Conference on Recommender systems, pp. 105-112, 2014.

[13] RVVSV Prasad,V.Valli Kumari, "Categorical Review of Recommender Systems," International Journal of Distributed and Parallel Systems ,Volume 3, Issue 5, 2012.

[14] Yehuda Koren, Robert Bell, Chris Volinsky, "Matrix factorization techniques for recommender systems," IEEE Computer Society Press Los Alamitos, Volume 42 Issue 8, pp. 30-37, 2009.

[15] Heng Luo, Changyong Niu, Ruimin Shen, Carsten Ullrich, "A collaborative filtering framework based on both local user similarity and global user similarity," Springer, Volume 72, Issue 3, pp. 231–245, 2008.

Author's Biographies

Ms.S.C.Lavanya completed her Bachelor of Engineering in 2009 and Master of Engineering in 2011. She is specialized in Computer Science and Engineering from Anna University Chennai. She is currently working as Assistant Professor in the department of Computer Science and Engineering at Dr. Mahalingam College of Engineering and Technology, Pollachi, Coimbatore, Tamil Nadu, India. She is a Life time member of ISTE, since 2011. Her research interests are Data Mining and Human Computer Interaction.

Ms.T.Gowrisankari completed her Bachelor of Engineering in 2013 and Master of Engineering in 2015. She is specialized in Computer Science and Engineering from Anna University Chennai. She is currently working as Assistant Professor in the department of Computer Science and Engineering at Dr. Mahalingam College of Engineering and Technology, Pollachi, Coimbatore, Tamil Nadu, India. She is a Life time member of ISTE, since 2015. Her research interests are Data Mining, IOT, Big data and Analytics and Human Computer Interaction.

Ms.B.Suganya completed her Bachelor of Engineering in 2013 and Master of Engineering in 2017. She is specialized in Computer Science and Engineering from Anna University Chennai. She is currently working as Assistant Professor in the department of Computer Science and Engineering at Dr. Mahalingam College of Engineering and Technology, Pollachi, Coimbatore, Tamil Nadu, India. She is a Life time member of ISTE, since 2017. Her research interests are Data Mining and Human Computer Interaction.