ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 6, June 2018

Grouping Like-Minded Users for Rating's Prediction: A Machine Learning Approach for Online Shopping System

R.K. Bharathi [1], Pavan Nagaraj Naik [2,] Mohana S D [3]

Associate Professor, Department of MCA, SJCE, JSS S&T University, Mysuru, Karnataka, India[1]

Department of MCA, SJCE, JSS S&T University, Mysuru, Karnataka, India[2]

Department of MCA, SJCE, JSS S&T University, Mysuru, Karnataka, India[3]

ABSTRACT: In the last few decades, several methodologies have been proposed by experts of research community in order to enhance online shopping system. Here we made an attempt to upgrading the existing system. Our recommender System aims at to predict future user ratings in order to estimate meaningful recommendations to a collection of users for items that users might like. Rating is ranking or grading the item based on user's satisfaction. Prediction is the act of saying what product the user might like in the future based on his previous ratings. Predicting the ratings is key task in recommender system. Recommender systems aim at finding items that are likely of interest to a user, by exploiting different types of information sources related to both the users and the items.

This paper presents algorithms for recommending items in online shopping system. Recommendation task is categorized as three steps: (1) grouping latent center of interest, (2) Clustering like-minded users, and (3) predict user item ratings. We use k-means clustering algorithm to create clusters of similar users. In order to predict item, we employ linear regression algorithm. Validate the result by employing the MAE and RMSE measures.

KEYWORDS: Collaborative Filtering, SVD++, Group recommendation and, Regression method.

I. INTRODUCTION

Due to huge number choices in internet, there is need to filter, prioritize and efficiently deliver relevant information to user, in order to solve this problem recommender system (RS) is used. Recommender systems have been widely used to provide users with high-quality personalized recommendations from a large volume of choices. Recommender systems have gained importance in recent years, by including in a variety of areas such as movies, music, news, books, research articles, search queries, social tags etc. RS produce content for users, by suggesting items that users might like. Recommender systems which are majorly used in the Media and Entertainment industries are: Collaborative Recommender system [3], Content-based recommender system, Demographic based recommender system, Utility based recommender system, Knowledge based recommender system and Hybrid recommender system. Two major factors such as high performance and simple requirements have made the Collaborative Filtering (CF) more popular. The basic principle of Collaborative Filtering is to create patterns based on user preferences and hence use the same patterns of preference to produce suggestions in later usage.

Nowadays, a large number of online shopping businesses are emerging rapidly and RS is considered predominantly for Item Recommendation also. For these emerging businesses, existing recommendation models usually suffer from the data-sparsity. Furthermore, a limitation is encountered here which results in non-performance of RS for new item purchases as they do not appear in the training data. Hence, to overcome this limitation, we introduce a Group Recommendation Methodology [5]. For more than one person use group recommend system. That is the classic group

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 6, June 2018

recommendation, the first step compute similar users in to identify the constraint on the number of recommendations that can be produce and maximize the users' satisfaction.
K-means Algorithm is used to create Like-Minded user groups where, an user u rating on a particular item i is used to infer based on stored ratings of the user u having similar interests. This will help to increase the recommendation accuracy and reduce the data dimensionality. Thus, we reduce the complexity and data sparsity problem. Building prediction for a group, leads to great improvement in the quality of recommendation.
The paper is structured as follows: The overview of the related work is presented in section 2 whereas; Section 3 presents approach to estimate ratings. A brief description about the dataset is presented in section 4 followed by the experimental results in Section 5. Conclusion and future works are listed in Section 6.

II. RELATED WORK

In order to understand existing system several research papers are studied, some of the research papers are discussed in this section. Number of approaches has been proposed for ratings prediction, including both Neighbourhood methods and Content-Based recommender systems. A survey on representative neighbourhood methods shows that neighbourhood recommender system can help more items for recommendations; accuracy is measured using RMSE measure. Similarly, propose Trust-based matrix factorization technique for recommendations, svd++ and Trust svd algorithms were used for recommendation. The survey continued to study the content-based collaborative filtering. And describe support-vector regression algorithm for rating's prediction and accuracy was calculated by MAE. An attempt is made on recommender system [13], Group recommendation method was discussed. Model Based, Merge Recommendations, and Predictions Aggregation algorithms were used for recommendation, whereas MovieLens-1M dataset was used to find the accuracy of algorithm. RMSE value is about 0.9872. Show that Social tagging system provides more external information to improve recommendation accuracy [2]. Recommender systems are designed to automatically generate personalized suggestions of products/services to customers. Product rating information can be used to generate preference-based recommendation.
Quite a few groups collaborative filtering model have been proposed to date. For example uses group collaborative filtering as a key feature in recommender system, GLER algorithm was used for movie recommendation on the Like-Minded user groups, to resolve the relative low accuracy and coverage issues of content based collaborative filtering. Similarly, propose a User–User CF algorithm and Item–Item CF algorithm for recommendations task [16].
Accuracy is a key feature of Recommendation system, and most of the existing system use MAE and RMSE measures. Existing approaches incorporates localized relationships into MF model by using co-occurrence properties for finding the set of neighbours. If the user only associates very few ratings for items, these approaches cannot obtain better performance. Therefore, we propose group based recommendation method product rating information can be used to generate preference-based recommendation [12]. Prediction of human behaviour on shopping [10] can be predicted by share their opinion like rating on the products [9]. Users rating could be predicted [4] by products availability/consuming. The major role of the online shopping is user usage durability; it means device usage diversity could be control by giving offers in online shopping on product this could be achieved to get good clustering of users in order to predict user needs and recommendation of products.

III. PROPOSED WORK

The latent factor is joint variation in response to unobserved variables, it describe the variability of observer among correlated to the low observer. The clustering technique is clustered the user item rating, and it recommend the item. By above likeminded peoples can be estimate using K- means algorithm methods.
We first introduce the concept of retrieve latent factor and group like-minded users, and to analyse the influence of rate for rating prediction based by real-world recorded data set. In the first step we group users based on interest centers and their opinions. Next, we apply clustering of like-minded users on the ratings prediction task in recommendation systems.
The main aim of the K-means clustering Algorithm is to create clusters of Like- Minded people such that it helps further to estimate future user-item ratings based on the stored ratings of the particular users having similar tastes.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 6, June 2018

Now, in order to estimate a new user-item rating for an existing user, the recommendation is done corresponding to the group that the user belongs to. Whereas, for a new user, the closest cluster model is used. In the closest cluster model we calculate the distance between the new user and the users from each cluster of similar interest. Predicting a user rating can be considered either as a regression or a classification problem or sometimes as both. If the ratings are represented in the form of classes or labels, it can be treated as a classification problem, whereas, the same is considered as a regression problem if the ratings are in the discrete form, i.e., if they are represented in numerical values. Illustrate five-star evaluation mechanism in grouping like-minded user rating, for prediction and specifically in online shopping recommendation to the system. So in this work, linear regression algorithm is adopted.

A. *CREATE GROUP*

In order to create clusters from group first we have to determine the user model in the group. Grouping will be done based on categories. Generally used to reduce the data dimensions or to retrieve the latent centres where the data is concentrated.

B. *CLUSTERING SIMILAR USERS*

There are many clustering algorithms presents. In this work we adopt K-Means Clustering algorithm to group the similar users. K-Means clustering is used for classifying/grouping items into k clusters (where k is the number of pre-chosen clusters). The clustering is done by minimizing the sum of squared distances (Euclidean distances) between items and the corresponding centroid. K-Means Algorithm is good compared to other clustering algorithm when the amount of data is expected to grow. K-Means algorithm is almost self-explanatory; the initial cluster means are randomly chosen values within the same range as the highest and lowest of the data values.
Mathematical form: K means

$$J = \sum_{j=1}^{K} \sum_{n \in S_j} |x_n - \mu|^2$$

Above equation is portioning the observation by means.

Where portioning N data points into K disjoint subsets Sj containing Nj data points, so as to be minimize the sum of squares criterion such that xn is vector and μj is centriods.

C. *PREDICT RATINGS*

Predicting a user rating could be treated as both a regression and a classification problem. In Online shopping system rating is five start evaluations so we considered prediction is regression issue. So in this work, linear regression algorithm is adopted. In this work predictor variables are continues valued and at the end of this work we need single predictor variable so we adopt straight line linear regression algorithm over different regression algorithm.

The relationship between one or more independent and predictor variable can be used to model using the linear regression algorithm. Predictor variable is user who purchased x item, and response variable is ratings (what we want to predict).
Mathematical form: Linear Regression.

$$b = \frac{\sum_{i=1} (x - \overline{x})(y_i - \overline{y})}{\sum_{i=0}^{|D|} (x - \overline{x})^2}$$

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 6, June 2018

Regression linear line is a response variable y and a single predictor variable x. The simplest form is linear function of y. Slope of linear regression can be calculated by using above formula.
Intercept of linear regression is:

$$a = \overline{y} - \overline{b.\ x}$$

And the regression value can be estimated using this method is:

$$w = a + b.\ x$$

IV. EXPERIMENTAL EVALUATION

A. *DATASET*

To analyze and evaluate the accuracy of the proposed algorithm, two different data sets are used.. First data set is Amazon 10 lakhs dataset collected by the Amazon dataset of ”clothing and shoes” through the Amazon product data web site. The dataset contains product ratings and metadata from Amazon, including 10, 48,576 ratings spanning May 1996 to July 2014. This data set consists of 10, 48,576 ratings [1] from 8, 46,468 users on 1, 19,373 items from 7 categories. Demographic information about the users is: user name, password, email, user id. The information about item in the dataset is: item id, category id, item name, image, and price. Second data set is Synthetic data set. it consists of 3300 ratings (1-5) from 500 users on 279 items from 7 categories. Demographic information about the users is: user name, password, email, user id. The information about item in the dataset is: item id, category id, item name, image, and price.

B. *ACCURACY MEASUREMENT*

In this section we conduct a series of experiments in order to investigate the effectiveness of our analysis work. To obtain the recommender system accuracy, we use two prevalent measures of accuracy, namely: Root Mean Squared Error and Mean Absolute Error. Two data tables are created for testing. First, rating table (pui) consists of all recommended items. Second, user purchase table (rui) consist of user purchase list.

Root Mean Squared Error (RMSE)

The quality of the predicted ratings was measured through the Root Mean Squared Error (RMSE). The metric compares each rating of user purchase table (rui) , expressed by a user u for an item i in the test set, with the rating table (pui) , predicted for the item i for the group g in which user u is. The formula is shown below:

$$RMSE = \text{sqrt}\left(\frac{\sum_{i=1}^{N} (rui - pui)^2}{N}\right)$$

Mean Absolute Error (MAE)

It is given by is a measure of two continuous variables; i.e., comparison between predicted and observed and assume that two variables of paired observation that express the same phenomenon.

$$MAE = \frac{\sum_{i=1}^{N} (rui - pui)}{N}$$

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 6, June 2018

C. *EXPERIMENTAL RESULTS*

To evaluate algorithm on dataset, we start by creating the input matrix, Thus, we create a user group { item_id, cat_id, user_id, cat_name, gp_count, rate_avg }, where ” item_id”, ” cat_id”, ” user_id”, and ” cat_name” represent respectively the item code, category id, user id and category name of the user x. ” gp_count” is the number of user who rated particular item belongs to particular category. ” rate_avg” is the average of ratings given to each category. Thus, grouping represents a user not only by his personal information but also by his interest to items. Then, we create “similarity” table. Which calculate count of users, who rated same rating for similar category? Using grouping and similarity table, we form cluster of like-minded users. Finally, we divide the training base into several dataset according to the users’ cluster, and we use linear regression algorithm to build a model for each group. Once learnt, we use these models to estimate ratings for users of the corresponding cluster.

In order to find accuracy of proposed algorithm, we create a user rate table user_rate = {user_id, item_id, cat_id, ratings }, where ” user_id”, ” item_id”, ” cat_id”, and ” ratings” represent respectively the user id, item id, category id and ratings given to item of the user x. compare this table with recommendation table created.

Evaluation metrics: We considered to evaluate predictive accuracy, by using mean absolute error (MAE) and root mean square error (RMSE). We also calculate precision and recall value of algorithm in order to find f measure. Tables 1 illustrate the accuracy of the linear regression algorithm on the dataset.

Table 1 Illustrate the accuracy of the linear regression algorithm on the dataset.

The RMSE value is difference between the predicted value and actual value. The actual value and forecasting value difference is MAE. Precision means how many selected elements are relevant. Recall means how many relevant elements are selected. F measure is calculating the effectiveness of the measurements

TABLE 1

RMSE	MAE	Precision	Recall	F measure
2.6177	0.7556	0.4133	0.8266	0.5510

V. CONCLUSION

In grouping like-minded users for ratings prediction work, we presented the linear regression algorithm for recommending items. Like-minded user groups in K-means algorithm creates based on the user-item ratings. Then, it uses those groups to build patterns and predict user-item ratings. Experimenting on the Amazon-10 lakhs dataset and synthetic 3300 dataset, Result shows that linear regression achieves better results than the TIMESVD++ algorithm and neighbourhood approach. This improvement manifests in the decrease of MAE and RMSE values of about 0.7556 and 2.6177 respectively.

VI. FUTURE WORK

The machine learning is the continuous process for adding new algorithm in order to build sophisticated model, above problem is in continuous process, it could not be stop until the human life existing in the world, but human needs augments products in his life. So online shopping could be more effective in satisfying customer, for this future recommendation based on related search things in the online can be a sentiment learning process on the customer products and its history of ratings and search. This sentiment learning process might be improved the online shopping.

ISSN(Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 6, June 2018

REFERENCES

[1] Singh, Abhinav, and Conrad S. Tucker. "A machine learning approach to product review disambiguation based on function, form and behavior classification." *Decision Support Systems*97 (2017): 81-91.
[2] Li, Jianguo, Yong Tang, and Jiemin Chen. "Leveraging tagging and rating for recommendation: RMF meets weighted diffusion on tripartite graphs." *Physica A: Statistical Mechanics and its Applications* 483 (2017): 398-411..
[3] Bobadilla, Jesús, Rodolfo Bojorque, Antonio Hernando, and Remigio Hurtado. "Recommender Systems Clustering using Bayesian non Negative Matrix Factorization." *IEEE Access*(2017)..
[4] Zafari, Farhad, Irene Moser, and R. Rahmani. "Proposing a highly accurate hybrid component-based factorised preference model in recommender systems." In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*. 2017.
[5] Jaffali, Soufiene, Salma Jamoussi, Abdelmajid Ben Hamadou, and Kamel Smaili. "Grouping Like-Minded Users for Ratings' Prediction." In *Intelligent Decision Technologies 2016*, pp. 3-14. Springer, Cham, 2016.
[6] Guo, Guibing, Jie Zhang, and Neil Yorke-Smith. "A novel recommendation model regularized with user trust and item ratings." *ieee transactions on knowledge and data engineering*28, no. 7 (2016): 1607-1620.
[7] Izadpanah, Saman. Utilizing one-class collaborative filtering using clustering and Top-K nearest neighbor. Lamar University-Beaumont, 2015.
[8] De Canio, F., M. Ieva, and C. Ziliani. "Device usage patterns and online shopping behaviour." In *Proceedings of the XII annual Conference Società Italiana Marketing*. 2015.
[9] Zhang, Yadong, and Du Zhang. "Automatically predicting the helpfulness of online reviews." In *Information Reuse and Integration (IRI), 2014 IEEE 15th International Conference on*, pp. 662-668. IEEE, 2014..
[10] Gupta, Mona, Happy Mittal, Parag Singla, and Amitabha Bagchi. "Characterizing comparison shopping behavior: A case study." In *Data Engineering Workshops (ICDEW), 2014 IEEE 30th International Conference on*, pp. 115-122. IEEE, 2014.
[11] Jeong, Hwa Young, Mohammad S. Obaidat, Neil Y. Yen, and James J. Jong Hyuk Park, eds. *Advances in Computer Science and Its Applications: Csa 2013*. Vol. 279. Springer Science & Business Media, 2013.
[12] Yoon, Victoria Y., R. Eric Hostler, Zhiling Guo, and Tor Guimaraes. "Assessing the moderating effect of consumer product knowledge and online shopping experience on using recommendation agents for customer loyalty." *Decision Support Systems* 55, no. 4 (2013): 883-893..
[13] Zhang, Zui, Hua Lin, Kun Liu, Dianshuang Wu, Guangquan Zhang, and Jie Lu. "A hybrid fuzzy-based personalized recommender system for telecom products/services." *Information Sciences* 235 (2013): 117-129.
[14] Boratto, Ludovico, and Salvatore Carta. "The rating prediction task in a group recommender system that automatically detects groups: architectures, algorithms, and performance evaluation." *Journal of Intelligent Information Systems* 45, no. 2 (2015): 221-245.
[15] Kaminskas, Marius, and Francesco Ricci. "Contextual music information retrieval and recommendation: State of the art and challenges." *Computer Science Review* 6, no. 2-3 (2012): 89-119.
[16] Boratto, Ludovico. "Group recommendation with automatic detection and classification of groups." PhD diss., Universita'degli Studi di Cagliari, 2012.
[17] Lops, Pasquale, Marco De Gemmis, and Giovanni Semeraro. "Content-based recommender systems: State of the art and trends." In *Recommender systems handbook*, pp. 73-105. Springer, Boston, MA, 2011.
[18] Sadikov, E., Madhavan, J., Wang, L. and Halevy, A., 2010, April. Clustering query refinements by user intent. In Proceedings of the 19th international conference on World wide web (pp. 841-850). ACM.
[19] Koren, Yehuda. "The bellkor solution to the netflix grand prize." *Netflix prize documentation* 81 (2009): 1-10.