Special Issue Published in International Journal of Trend in Research and Development (IJTRD), ISSN: 2394-9333, www.ijtrd.com

A Conceptual Study of Recommender Systems

M.Suguna
Assistant Professor, Department of Computer Science, Periyar University, St. Joseph's College of Arts and Science for Women, Hosur, India

Abstract - Data mining is one of the inter disciplinary fields on the research aspect and particularly research on web mining is the most important part. A prediction technique in data mining plays a major role in the field of recommendation system. The recommender system is to recommend good items to the users. The core objective of this paper is to discuss basic concepts of recommender system and also to describe the basic phases of the recommender system.

Keywords — Data mining, Prediction, Recommendation, Collaborative Filtering (CF), normalization, similarity measures, imputation methods, evaluation metrics.

I. INTRODUCTION

In the present scenario, information technology penetrates into every part of human life. With the enormous amount of data stored in files, databases and other repositories, it is increasingly important to develop powerful means for analysis, interpretation of such data and for the extraction of interesting knowledge that could help in decision-making. Data mining is a multidisciplinary and emerging field.

Currently, consumers are faced with millions of goods and services when shopping on-line. Recommender systems (RSs) help consumers by making product recommendations during live customer transactions. Technically, recommender system has their origins in different fields such as Information Retrieval (IR), text classification, machine learning and Decision Support Systems (DSS).

A collaborative filtering (CF) approach is commonly used, in which products are recommended based on the opinions of other customers. Collaborative recommender systems may employ data mining or statistical techniques to search for similarities among customer preferences. Collaborative filtering technique is based on K-Nearest Neighbor classification algorithm.

Rest of the paper is organized as Section 2 reviews related works in this field, Section 3 presents the basic elements, user-item matrix, techniques and issues of recommender system, Section 4 discusses the Collaborative Filtering technique, Section 5 presents the phases of Recommender system. Finally, Section 6 concludes our discussion.

II. RELATED WORKS

Prediction has become one of the important roles in data mining tasks as well as recommender system. To give accurate prediction, data are pre-processed and then unknown ratings are predicted and good items are recommended.

Goldberg, K. et al., [1] proposed Eigentaste CF algorithm. It normalized each rating by subtracting its mean rating over all users, and then dividing by its standard deviation. It also applied principal component analysis for dimensionality reduction.

Symeonidis, P. et al. [2] used a model-based approach based on latent semantic indexing (LSI), combining simultaneously effectiveness and efficiency of CF algorithms. First it applied the Singular Value Decomposition to reduce the dimensionality of user-item matrix. Then it used a ranking criterion denoted as highest predicted rated item recommendation (HPR) that uses the predicted values of a user for each item.

Ghazanfar M A et al. [3] proposed nearly 17 approaches to approximate the missing values in the user-item rating matrix. The main approaches are based on various classifiers such as k-nearest neighbor, Naïve bayes and support vector machine.

Babu, M S P. et al. [4] implemented the user based collaborative filtering algorithm using the Pearson correlation similarity measure in Java and executed the results. It used the Jester data set and stored it in database MySQL

Nilashi M et al. [5] presented the basic concepts such as data format, CF techniques, CF tasks such as prediction and recommendation, similarity measures, evaluation metrics used in CF system.

Bagchi, S. et al. [6] analyzed performance and quality aspects of different similarity measures such as Euclidean distance similarity, City-block similarity, Un centered cosine similarity, Pearson correlation similarity, Spearman correlation similarity, Tanimoto coefficient similarity, Log likelihood similarity used in collaborative filtering.

III. RECOMMENDER SYSTEM

Recommender System (RS) is a software tool and techniques providing suggestions for items a user may wish to utilize [5].

A. Basic Elements

Generally, three kinds of elements can used by recommender systems are items, users and transactions. These elements can be used in data construction.

Items T = {t1, t2,…,tn} are the products in the recommender system for suggesting to the user.The value of an item is considered positive when the item is useful for the user and the value is negative when the item is inappropriate due to the wrong decision made in selection process.

Elements of U = {u1, u2,…,un} comprises of all the users that have browsed items or contributed to the item ratings in the sites. In collaborative filtering system, users are modelled as a simple list containing the ratings provided by the user for some items.

The rating is in fact the most popular form of transaction data collected by a recommender system which may be explicitly or implicitly. In the explicit rating, the user is asked to provide about an item on a rating scale.

There are different kinds of ratings that could be adopted for recommendation [7] as follows:

1. **Numerical ratings** represented by a number from either a discrete or a continuous rating scale. Discrete rating scale are ratings on a scale from zero to five stars while continuous rating scale could be a slider set by a user and translated to a real value.
2. **Binary ratings** scale allowing users to assign items to two different classes (like/dislike)
3. **Ordinal ratings** such as "strongly agree, agree, neutral, disagree, strongly disagree"
4. **Unary rating** by contrast, allows users to assign items only to a single class, which is positive in most cases and a prominent example is the Facebook's "Like"-button.

B. User-Item Matrix

User-item matrix is a matrix of customers against products that have components as the explicit ratings of customers to products (user to item). Some of the user-matrix cells are not loaded, as there are products that are not rated by any user.

For M items and K users, the user profiles are represented in a K×M user-item matrix X. Each element $x_{k,m} = r$ indicates that user k rated item m by r, where r €{1, …, |r|}. If the item has been rated and $X_{k,m}= 0$; means that the rating is unknown.

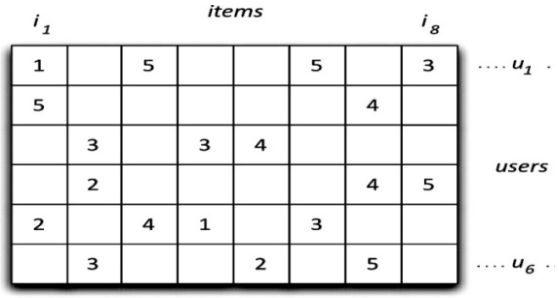

Figure 1: User-Item Matrix

The user-item matrix containing the ratings by six users for eight items is shown in figure 1.

The rows of the User-Item Matrix denote the users and the columns denote the items and the values contained in a matrix are ratings in the range 1 to 5. The blank space denotes the users have not rated the items.

C. Recommendation Techniques

There are six different classes of recommendation approaches [8]. They are as follows:

1. Content Based Recommender System

Content recommender systems try to suggest products that are similar to the ones that the user liked in the past. The likeness of items is determined depending on the traits associated with the compared items.

2. Collaborative Filtering System

A collaborative filtering (CF) approach is commonly used, in which products are recommended based on the opinions of other customers.

3. Demographic Recommender System

This type of system suggests items depending on the user demographic profile. For example, customers are dispatched to specific websites depending on their language or nation.

4. Knowledge Based System

Knowledge-Based (KB) systems recommend items based on particular domain knowledge about how certain item features fulfill users' needs and preferences and, eventually, how the item is useful for the user.

5. Community Based System

This kind of system works on the preferences of the user's friends to recommend items.

6. Hybrid Recommender System

Hybrid recommender system can be obtained from a combination of mentioned techniques by combining two or more techniques that tries to alleviate disadvantages of them.

D. Issues of Recommender System

The main issues and challenges in recommender systems [9] are as follows.

1. Cold Start Problem

This problem occurs when new users enter the system or new items are added to the catalogue. In such cases, neither the taste of the new users can be predicted nor can the new items be rated or purchased by the users leading to less accurate recommendations.

2. Synonymy

Synonymy arises when an item is represented with two or more different names or entries having similar meanings [10].

3. Shilling Attacks

What happens if a malicious user or competitor enters into a system and starts giving false ratings on some items either to increase the item popularity or to de-crease its popularity [11]. Such attacks can break the trust on the recommender system as well as decrease the performance and quality of recommenders.

4. Privacy

Feeding personal information to the recommender systems results in better recommendation services but may lead to issues of data privacy and security.

5. Sparsity

Sparsity is a problem common to most recommender systems due to the fact that users typically rate only a small proportion of the available items.

6. Scalability

The rate of growth of nearest-neighbour algorithms shows a linear relation with number of items and number of users. It becomes difficult for a typical recommender to process such large-scale data.

IV. COLLABORATIVE FILTERING RECOMMENDATION TECHNIQUE

All collaborative filtering methods share a capability to utilize the past ratings of users in order to predict or recommend new content that an individual user will like.

Collaborative recommender systems may employ data mining or statistical techniques to search for similarities among customer preferences.

A collaborative recommender system works by finding a set of customers, referred to as neighbors, that have a history of agreeing with the target customer (such as, they tend to buy similar sets of products, or give similar ratings for certain products)

As with classification systems, recommender systems can make two types of errors: false negatives and false positives. Here, **false negatives** are products that the system fails to recommend, although the consumer would like them. **False positives** are products that are recommended, but which the consumer does not like.

Two basic variants of collaborative filtering approach can be classified as **user-based and item-based.**

According to Breeze et al., [9], methods for collaborative recommendations can be classified into two groups: memory-based and model-based.

1. **Memory-based methods** [12] are heuristics that make ratings predictions depending on the whole collection of items formerly rated by users. These techniques require all ratings, items and users to be maintained in memory..
2. **Model-based methods** [1] use the group selection of ratings to learn a model, which is then used to make rating predictions. These techniques regularly make a concise of ratings patterns off-line.

IV. PHASES OF RECOMMENDER SYSTEM

Usually the recommender system recommends good items by going through three phases. Each phase uses different data mining, supervised machine learning algorithms.

The raw data is processed in the sequence as shown in figure 2 which yields good recommendation of items

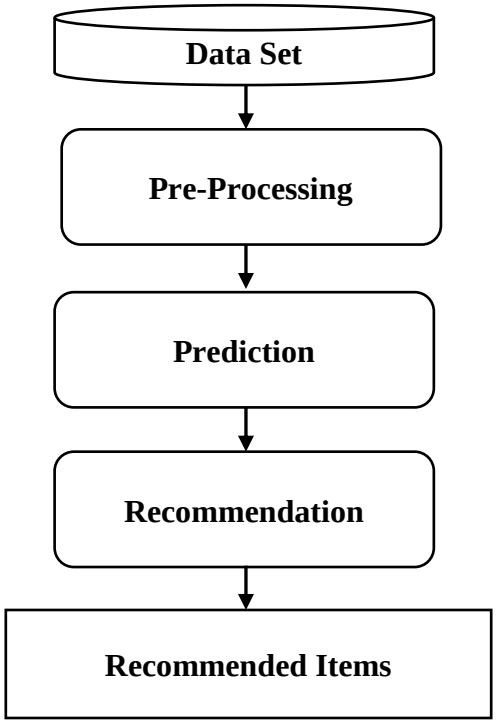

Figure 2: Phases of Recommender System

A. *Pre-processing*

Preprocessing of raw data can significantly reduce the prediction error of ratings and can increase the quality of recommendation of good items.

Data Cleaning is the process of removing missing values and various noises.

The raw data is converted into real ratings matrix to clean the data. The cleaning process of missing values neglects the unrated entries and considers only rated entries to form real ratings matrix.

Normalization is the process where the attribute data are scaled so as to fall within a small specified range,

such as -1.0 to 1.0, or 0.0 to 1.0. Normalization may improve the accuracy and efficiency of mining algorithms involving distance measurements.

Different normalization methods are mean normalization, subtractive and multiplicative normalization and z-score normalization. Two of the most popular rating normalization schemes that have been proposed to convert individual ratings to a more universal scale are mean centering and Z-score.

The real ratings matrix is normalized to overcome user biases.

B. *Prediction*

Generally, data mining tasks are classified into two categories:

1. **Descriptive:** The descriptive mining describes the general properties of the existing data.
2. **Predictive:** Predictive mining tasks perform inference on the current data in order to make predictions.

Prediction is the process to predict the numeric values of the attributes (items). Prediction is one of the supervised machine learning algorithms.

The techniques mainly used for prediction are Regression analysis and some classification techniques such as back propagation, support vector machines, and k-nearest-neighbor classifiers.

1. Regression Analysis

Regression analysis is a statistical methodology that is most often used for numeric prediction. Regression analysis can be used to find the relationship between one or more independent or predictor variables and a dependent or response variable (which is continuous-valued).

Linear regression involves finding the "best" line to fit two attributes (or variables), so that one attribute can be used to predict the other**. Multiple linear regressions** is an extension of linear regression, where more than two attributes are involved and the data are fit to a multidimensional surface.

Other regression-based methods are generalized linear models, poisson regression, log-linear models, and regression trees

2. Classification Techniques for Prediction

Classification techniques such as back propagation, support vector machines, and k-nearest-neighbor classifiers can be adapted for prediction of numerical attributes. As the KNN is a basis for collaborative filtering algorithm and is discussed as follows.

The K-Nearest Neighbor classifier can be used to predict the missing or unavailable numerical data values for the test data by selecting the k nearest neighbors from the training data using the similarity measures such as Pearson correlation, Cosine measure and Euclidean distance and also aggregation functions.

The metrics used to assess the performance of predictor are as follows:

1. Mean Absolute Error (MAE)

The MAE is determined as the average absolute deviation between predicted ratings and true ratings. It is measured by Equation 5.1

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|OR_i - PR_i| \qquad (5.1)$$

where OR_i denotes the true ratings of an item i and PR_i denotes the predicted ratings of an item i.

2. **Mean Squared Error (MSE)**

The MSE exaggerates the presence of outliers that is it does not consider the small prediction errors. It is measured using the formula in Equation 5.2.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(OR_i - PR_i)^2 \qquad (5.2)$$

3. **Root mean squared error (RMSE)**

The RSME is the square root of MSE. Its formula is given in Equation 5.3.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(OR_i - PR_i)^2} \qquad (5.3)$$

C. Recommendation

Recommendation is the process to recommend items to the new user depending on the predicted ratings.

Usually the items whose predicted ratings are of good ratings are recommended.

Researchers have used different recommendation methods such as random items, popular items, user based and item based collaborative filtering algorithm.

The TOP-N recommendation method is the basis for all the recommendation algorithms. The recommendation process recommends good items using the prediction-threshold based Best-N recommendation method. This method recommends only the items whose predicted ratings are greater than particular threshold value (Eg. +5 for Jester data set).

The quality of the recommendation is judged using the k-fold cross validation technique by representing the precision and recall.

Good recommendation system produces higher precision and recall. Precision denotes recommendation accuracy and recall denotes filtering accuracy.

1. **Precision**

Precision is defined as the ratio of relevant items selected to the number of items selected.

2. **Recall**

Recall is defined as the ratio of relevant items selected to the total number of relevant items available.

CONCLUSION

This paper introduces the recommender system. It then explains the basic elements and data format used by the recommender system. It lists out the issues of recommender system. Then this paper specifies various recommendation techniques and describes collaborative filtering technique in detail. Next it explains the three working phases of recommender system such as Preprocessing, prediction and recommendation. It defines the metrics used to evaluate the prediction of ratings such as MAE, MSE, RMSE. The quality of recommendation is judged using the evaluation metrics such as Precision and Recall. In Future, this paper can be used by beginners of research to know about Recommender system and also to also to make findings in this field.

References

[1]. Goldberg, P. K., Roeder, T., Gupta, D.,and Perkins, C., "Eigentaste: A constant time collaborative filtering algorithm", Information Retrieval, Vol 4, No 2, pp. 133-151, 2001.

[2]. Symeonidis, P., Nanopoulos, A., Papadopoulos, A. N., and Manolopoulos, Y., "Collaborative recommender systems: Combining effectiveness and efficiency", Expert Systems with Applications, Vol. 34, No. 4, pp. 2995-3013, 2008.

[3]. Ghazanfar, M. A., &Prügel-Bennett, A. 2013. The Advantage of Careful Imputation Sources in Sparse Data-Environment of Recommender Systems: Generating Improved SVD-based Recommendations. Informatica (Slovenia),37(1), 61-92.

[4]. Babu, M. S. P., AND Kumar, B. R. S., "An Implementation of the User-based Collaborative Filtering Algorithm", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol.2, No.3, pp.1283-1286, 2011.

[5]. Nilashi.M.,Bagherifard.K., Ibrahim.O., Alizadeh.H., Nojeem.L.A and Roozegar.N., 2013.Collaborative Filtering Recommender Systems, Research Journal of Applied Sciences, Engineering and Technology, (vol.5,no.16,pp. 4168-4182)

[6]. Bagchi, S., "Performance and Quality Assessment ofSimilarity Measures in Collaborative Filtering usingMahout", Procedia Computer Science, No. 50, pp.229-234,2015.

[7]. Schafer, J. B., Frankowski, D., Herlocker, J., and Sen, S., "Collaborative Filtering Recommender Systems", In: Brusilovsky, P., A. Kobsa and W. Nejdl (Eds.), the Adaptive Web. Springer Berlin, Heidelberg, pp. 291-324, 2007a.

[8]. Burke, R. D., "Hybrid web recommender systems", Lect. Notes Comput. Sc., 4321, pp. 377-408, 2007.

[9]. Khusro, S., Ali, Z., Ullah, I., "Recommender systems: Issues, challenges, and research opportunities" Information Science and Applications (ICISA) 2016, 2016 - Springer

[10]. Su, X., Khoshgoftaar, T.M., "A survey of collaborative filtering techniques", Advances in Artificial Intelligence 2009, 4 (2009).

[11]. Gunes, I., Kaleli, C., Bilge, A., Polat, H.., "Shilling attacks against recommender systems: a comprehensive survey", Artificial Intelligence Review 42, 767–799 (2014).

[12]. Breese, J. .S., Heckerman, D., and Kadie, C. "Empirical analysis of predictive algorithms for collaborative filtering", Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, pp. 43-52, 1999.