Recommender systems: An overview of different approaches to recommendations

Conference Paper · March 2017

DOI: 10.1109/ICIIECS.2017.8276172

Saurabh Dongare
Liverpool John Moores University

2017 International Conference on Innovations in information Embedded and Communication Systems (ICIIECS)

Recommender Systems: An overview of different approaches to recommendations

Kunal Shah[1], Akshaykumar Salunke[2], Saurabh Dongare[3], Kisandas Antala[4]
[1, 2, 3, 4]Dept. Computer Science,
Sinhgad Institute of Technology,
Lonavala, India.

Abstract— **This paper presents an overview of the field of recommender systems and describes the present generation of recommendation methods. Recommender systems or recommendation systems (RSs) are a subset of information filtering system and are software tools and techniques providing suggestions to the user according to their need. Many popular E-commerce sites widely use RSs to recommend news, music, research articles, books, and product items. Recommendation systems use personal, implicit and local information from the Internet. This paper attempts to describe various limitations of recommendation methods and their advantages.**

Index Terms— **Recommender systems, information retrieval system, recommendations, collaborative filtering, content filtering, hybrid filtering.**

I. Introduction

Recommender systems became an active area of research since the appearance of the first papers on collaborative filtering in the mid-1990s. In recent years, many websites are widely using Recommender system. "You may know this," "other products you may like" ,"customers who bought this item also bought," "you may like this." Everyone has seen these suggestions when browsing the web, be it on Facebook or Amazon or some other platform. As these Web sites began to develop, a pressing need emerged for providing recommendations compiled from filtering the whole range of available options. The drastic growth in data and variety of information available on the Web and the rapid introduction of new E-business services (product buying, product comparison, auctioning, etc.) often overwhelmed users, leading them to make inappropriate decisions. From the immense variety of items (services and products) that these Web sites were offering, users were finding it very critical to make the most appropriate choices.

A. Recommender System Background

RSs are basically software tools and techniques for information retrieval and filtering that aims to provide meaningful and effective item recommendations to the active user [2]. These suggestions also called as Recommendations relate to various decision-making process such as which news article to read, what products (items) to buy or which song to listen. "Item" is a general term that is used to denote what the RSs recommends to the users. Recommendations are of two types, personalized and non-personalized. In personalized recommendations, different users or user groups receive different suggestions. In non-personalized recommendations, all the users get same suggestions. RSs are generically classified into the following categories, considering how recommendations are made: Collaborative recommendations, Content-based recommendations, and Hybrid approaches. The following sections describe the most popular techniques used now-a- days for building RSs such as Collaborative filtering, Content-based and Hybrid approach.

II. COLLABORATIVE FILTERING

The term "collaborative filtering (CF)" was coined by Goldberg et al., in 1992 who proposed that the information filtering process becomes more effective when humans are involved. [3]. In the collaborative filtering method, recommendations for each user are generated by making comparisons with the liking for one alternative over another of other users who have qualified the product similarly to the active user. CF are based on the idea that people who agree with the evaluation of items in the past are likely to agree again in future. CF methods are grouped into two general methods neighborhood based and model-based [1].

A. Neighborhood Based Method

In neighborhood based (memory-based or heuristic-based) CF, the user-item ratings stored in the memory are used directly to predict classifications for new items. which can be done in two ways, first user-based and second item-based recommendation. The User-based approach evaluates the interest of a user u for an item i by taking into consideration the ratings for this item by other users, called as neighbors that have similar rating patterns. The item-based approach predicts the rating for an item i of a user u based on the ratings of u for items similar to i. In item-based approach, the similarity between two items is calculated by the ratings provided by other users of the system in a similar way.

The main advantages of neighborhood-based methods are:-

- Simplicity: Relatively simple to implement.
- Justifiability: Provide a brief but comprehensive and easily understandable justification for the computed predictions.
- Efficiency: Do not require costly training phases, which need to be carried out at frequent intervals in large commercial applications.

978-1-5090-3294-5/17/$31.00 ©2017 IEEE

- Stability: Little influence by the constant addition of items, users, and ratings, which are typically observed in major E-commercial applications.

Formal Definition of problem: Set of users in the system are denoted by U, the set of items by I, and the set of ratings by R in the system, and S is the set of possible values for a rating (e.g. S = [1 through 5] or S =like, dislike). The notation u_i denotes the subset of users that have rated an item i. I_u represents the subset of items rated by user u. Two important problems associated with RS are the *best item* and *top-N recommendation* problems. The first problem consists of finding, the new item i € I or Iu for a particular user u, in which user is most likely to be interested. This task of finding such item for the user is modeled as a regression or classification problem. Where the aim is to find a function $f : U \times I \rightarrow S$ that predicts the rating $f(u, i)$ of a user u for a new item i. The obtained function then recommends an item k to the u_k (active user) for which the predicted rating has the highest value:

$$k = \operatorname{argmax} f(u_k, j)$$

Accuracy is used to evaluate models performance of the recommendation method. Usually, the rating set R is split into a training set R_{tr} used to learn f , and a test set R_{te} used to evaluate the prediction accuracy. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are two of the most common metrics used to measure accuracy. MAE assess the average magnitude of the errors in a generated set of predictions, without taking into consideration their direction. RMSE is a quadratic scoring rule that also assesses the average magnitude of the error. It's the square root of the average of squared differences between prediction and actual observation. Consider a situation where only the items purchased by each user are known to the RS, measuring the rating prediction accuracy is not possible. In such cases, the problem of finding the best item is usually modeled into the task of predicting recommendations to an active user u_k a list L(u_k) containing N items likely to interest him or her [8, 9]. The quality of this method can be measured by splitting the items of I into a set I_{tr}, used to learn L, and a test set I_{te}. Consider T(u) is the small subset of test items that a user u found relevant. So, the performance of the method is then computed using the measures of precision and recall:

$$\text{Precision } (\mathrm{L}) = \frac{1}{\mathrm{u}} \sum \left| \mathrm{L}(\mathrm{u}) \cap \mathrm{T}(\mathrm{u}) \right| / \mathrm{L}(\mathrm{u})$$

$$\text{Recall } (\mathrm{L}) = \frac{1}{\mathrm{u}} \sum \left| \mathrm{L}(\mathrm{u}) \cap \mathrm{T}(\mathrm{u}) \right| / \mathrm{L}(\mathrm{u})$$

B. Model-Based Recommendation Methods

Neighborhood based systems, which uses the stored ratings directly into the prediction whereas model based approaches use these ratings to acquire knowledge and learn a predictive model. The idea used in model based systems is to imitate and model the interactions of user-item with factors representing the hidden characteristics of the users and the items in the system, like the preference category of users and the category class of items. Then the model is trained using the available data (train dataset) and trained model is used to predict ratings of users for new items. There are many Model-based approaches for the task of recommending items. Which include techniques such as Bayesian Clustering [4], Latent Semantic Analysis [5], Support Vector Machines [6], and Singular Value Decomposition [7].

Advantages of collaborative filtering technique are listed below:

- Implementation of RS using Memory-Based Collaborative filtering technique is easy.
- Additional of new data in an incremental manner is easy in Memory-Based Collaborative filtering.
- Improved prediction performance by a Model Based Collaborative filtering technique.

The disadvantage of collaborative filtering technique:

- Cold Start Problem: In the case of the new user the system either doesn't know what to recommend or has very poor performance.
- Scalability: The CF technique generates recommendations over billions of users and products, which require a significantly huge amount of computational power.
- Sparsity: Only small subsets of the items are rated by the users from the available dataset of items. Hence very few ratings are available to generate a recommendation which leads in poor performance.

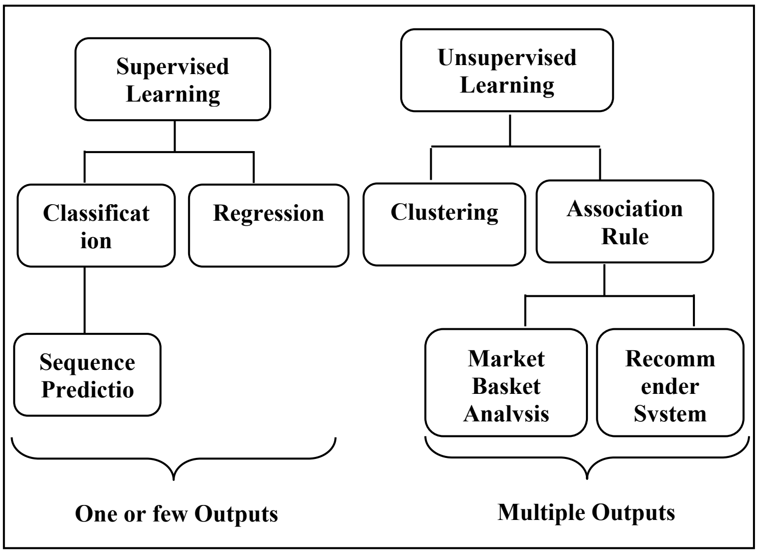

Fig 1. RSs in Machine Learning

The above figure clearly shows that recommender system comes under the Unsupervised Learning of the Machine Learning paradigm. In unsupervised learning the data available to us is not labeled so the hidden associations and cluster formation is revealed by performing unsupervised learning using algorithms that comes under that paradigm. Different approaches to recommendations are nothing but algorithms that comes under sub category Association rule mining.

TABLE I. TABLE TYPE STYLES

Paper Title	Approach to recommendation	Conclusion
Recommender Systems Handbook. Springer.	CB, Neighborhood-based, CF, Multi-Criteria Recommender, Robust CF.	The hybrid robust filtering methods are preferable over unique approaches.
A Case Based Recommendation Approach for Market Basket Data, IEEE intelligent systems.	CB, CF, Case-based reasoning (CBR).	CBR a new approach in hybrid filtering is considered as the preferable method for recommendation assuming transactions as the case.
Music Recommendation System.	CB, Low-level Descriptors, Correlation Analysis, Feature Vector Effectiveness, Collaborative filtering.	Optimization of both the feature vector and classification algorithm is essential.
Modeling relationships at multiple scales to improve the accuracy of large recommender systems.	CF,CB, Neighborhood-aware factorization	Designed new CF methods based on models that try hard to minimize quadratic errors, and demonstrated high performance on a large, real-world dataset. Hence they find CF to be more preferable.
Item-based top-N recommendation algorithms.	CF-based top-N recommender Systems: user-based, model-based. Item-based top-N recommendation algorithms.	Conditional probability-based item similarity scheme and higher-order item-based models provide reasonably accurate recommendation and are better than traditional user-based CF techniques. Implemented scheme is independent of the size of the user–item matrix.
kNN versus SVM in the collaborative filtering framework.	User profiling, collaborative filtering, Support Vector Machine, k-Nearest Neighbor	kNN is dominant on datasets with relatively low sparsity. On datasets with high to extremely high level of sparsity. In case of extreme sparsity, it is best to use a model-based approach, such as SVM classifier or SVM regression

III. CONTENT-BASED FILTERING

Content-based filtering, sometimes also referred to as cognitive filtering, recommends the items based on a comparison between the content of the items and a user profile data. The content of each item is represented as a set of descriptors, terms or feature vector. This descriptor can be a genre for movie item or can be frequent terms that occur in a document for the document as an item. The content based filter analyzes set of descriptors feed to it as input for a particular item which is previously rated by the user. This filter then constructs a model of users interest which generates recommendations [10]. For example, if a user likes a web page with the words "Camera," "Lens" and "Tripod" the Content-based filtering will recommend pages related to the electronics world. The item description and a profile of the user's orientation are important in Content-based filtering. CB filters try to generate recommendations based on similarity count. The top, matching items are recommended by comparing more than one candidate items with items that are already rated by the user.

Advantages of Content-based filtering:-

- Content-based recommender system provides user independence through exclusive ratings which are used by the user to build their profile.
- Provides transparency to the active user by providing an explanation of working of CB filter.
- CB filter is good to recommend items that are not yet rated or viewed by any user. This will be advantageous for a new user.

Limitations of Content-based filtering :-

- In CB Filters, it is very difficult to generate characteristics of an item.
- CB Filters suffer from an over-specialization problem because it advocates the same types of items.
- It is more difficult to get feedback from users in CBF because users normally do not sort the items (as in CF) and therefore, it is not possible to determine if the recommendation is correct.

IV. HYBRID APPROACH TO RECOMMENDATIONS

A wide and diverse variety of techniques are proposed for generating recommendations which include collaborative, content based, knowledge based and other techniques. These methods are blended in hybrid recommenders to improve performance. Collaborative filtering and Content-based filtering approaches are extensively used in information filtering application. Commonly, collaborative filtering is integrated with other techniques to nullify the ramp-up problem. Hybrid approaches can be put into action in various ways:

- Individual implementation of collaborative and content-based methods and aggregation of their predictions to generate recommendations.
- Integration of some pro characteristics from content-based methods into a collaborative approach,
- Integration of some pro characteristics from collaborative approach methods into a content-based approach,
- A generic consolidative model that is the assimilation of both content-based and collaborative characteristics.

A. Weighted

In the weighted hybrid recommendation score or weight of a recommended item is calculated from the results of all available recommendation techniques implemented in the system. Recommended components which have different scores are combined statistically. Additive aggregation is implemented to get normalized scores.

B. Switching

In this approach, out available recommendation components at the disposal, the system chooses one particular component and applies the picked one which best suits the purpose. The system has a criterion function to switch between recommendation techniques.

C. Mixed

Different recommendation approaches provide different recommendations that will be introduced together. This hybrid system is based on merging and presenting multiple rated lists into a single rated list. This approach avoids the "new item" start-up problem.

D. Feature combination

The hybrid system is divided or separated into two parts: contributing and actual recommender. These two co-exist in the system. The actual recommender depends upon the results or data output of the contributing recommender. Feature combination hybrid technique lets the system consider collaborative recommender output data without depending on it exclusively, which decreases the sensitivity of the system.

E. Feature Augmentations

Feature Augmentation hybrid is similar to the feature combination hybrids, but the difference is contributor gives interesting characteristic. Feature Augmentation hybrid is more elastic than feature combination method. Feature Augmentation found to make a significant contribution to the quality value of recommendations.

V. Conclusion

Recommender systems have become popular. RSs are widely utilized and deployed in a variety of areas. RSs provide recommendations based on users interaction with the system. Popular approaches to recommendations are content-based, collaborative and Hybrid. An overview of all these approaches shows that they have both advantages and disadvantages. CF system does not need content information about items and users to be machine recognizable. CF approach utilizes only ratings and there is no need of any additional information about users or items. But the principal disadvantage is that CF systems cannot generate recommendations if there are no ratings available. In the starting phase of development of such intelligent systems, the correctness and validity of the information that is served to the user were of concern. The memory based approach was unable to provide significant accuracy, but it can be used in simpler scenarios. Model based approach was developed which suffered from the problem of cold start to improve the quality of collaborative filters. To deal with the problem of cold start, Hybrid Recommender Systems as a result of extensive research were designed. Hybrid systems provide a notable enhancement in accuracy, precision, and recall matrices.

References

[1] Recommender Systems Handbook. Francesco Ricci,Lior Rokach, Bracha Shapira, Paul B. Kantor. Springer 2010.

[2] A Case-Based Recommendation Approach for Market Basket Data Anna Gatzioura and Miquel Snchez-Marr IEEE INTELLIGENT SYSTEMS 2015.

[3] Using collaborative filtering to weave an information Tapestry D. Goldberg, D. Nichols, B. M. Oki, and D. Terry,Communications of the ACM, vol. 35, no. 12, pp. 6170, 1992.

[4] Breese, J.S., Heckerman, D., Kadie, C.: Empirical analysis of predictive algorithms for collaborative filtering. In: Proc. of the 14th Annual Conf. on Uncertainty in Artificial Intelligence, pp. 4352. Morgan Kaufmann (1998).

[5] Hofmann, Collaborative filtering via Gaussian probabilistic latent semantic analysis. In: SIGIR 03: Proc. of the 26th Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval, pp. 259266. ACM, New York, NY, USA (2003).

[6] Grcar, M., Fortuna, B., Mladenic, D., Grobelnik, M.: k-NN versus SVM in the collaborative filtering framework. Data Science and Classification pp. 251260 (2006).

[7] Bell, R., Koren, Y., Volinsky, C.: Modeling relationships at multiple scales to improve the accuracy of large recommender systems. In: KDD 07: Proc. of the 13th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, pp. 95104. ACM, New York, NY, USA (2007).

[8] Deshpande, M., Karypis, G.: Item-based top-N recommendation algorithms. ACM Transaction on Information Systems 22(1), 143177 (2004).

[9] Last.fm: Music recommendation service (2009). http://www.last.fm

[10] Mladenic, D.: Text-learning and Related Intelligent Agents: A Survey. IEEE Intelligent Systems14(4), 44–54 (1999).

[11] Karim, J., 2014, Hybrid Systems for Personalized Recommendations,
Research Challenges in Information Science (RCIS), 2014 IEEE Eighth
International Conference, May 2014

[12] G. Linden, B. Smith and J. York, "Amazon.com recommendations item-to-item collaborative filtering." In IEEE Internet Computing, Volume 7,Issue 1, January 2003, pp.76-80, 2003