ISSN: 2321-7782 (Online)
Volume 2, Issue 11, November 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Review Paper on Clustering Based Collaborative Filtering

Sheetal Thokal
Department of Computer Engineering
JSPM's Bhivarabai Sawant Institute of Technology and Research, Wagholi, Pune, India.

Prof. Vrunda Bhusari
Department of Computer Engineering
JSPM's Bhivarabai Sawant Institute of Technology and Research, Wagholi, Pune, India.

Abstract: ***This short paper reports review on collaborative filtering which uses clustering algorithms. Nowadays Service relevant data become too big to be effectively processed by traditional approaches, so one solution to this challenge is Clustering Based Collaborative Filtering. This approach recruits similar services in the same cluster to recommend services collaboratively.***

Keywords: Big data, Clustering, Collaborative Filtering, Portioning, Recommender systems

I. INTRODUCTION

Big Data concerns large-volume, complex, growing data sets with multiple, autonomous sources. In Big Data applications data collection has grown tremendously and it is beyond the ability of commonly used software to capture, manage, and process that data. The most fundamental challenge for the Big Data applications is to explore the large volumes of data and extract useful information or knowledge for future actions. Recommender systems (RSs) are techniques and intelligent applications to assist users in a decision making process where they want to choose some items among set of alternative products or services. RSs encounter two main challenges for big data application: 1) to make decision within acceptable time; and 2) to generate ideal recommendations from so many services.

Clustering are such techniques that can reduce the data size by a large factor by grouping similar services together. Therefore, Clustering-based Collaborative Filtering approach (ClubCF), which consists of two stages: clustering and collaborative filtering is proposed. Clustering is a preprocessing step to separate big data into manageable parts. A cluster contains some similar services. As the number of services in a cluster is much less than the total number of services, the computation time of CF algorithm can be reduced significantly. Besides, since the ratings of similar services within a cluster are more relevant than that of dissimilar services, the recommendation accuracy based on users ratings may be enhanced.

II. PRAPOSED SYSTEM

Clustering based collaborative filtering approach contains tow modules. First Clustering, in this services are clustered depend on similarity in Description, Functionality & Characteristics respectively. Second Collaborative Filtering, in this, first rating similarity is computed & then predicted rating is given to the clustered services.

a) Clustering

In step clustering first stem words are recognized by using Porter Stemmer Algorithm. Then similarity in services based on Description & Functionality is Computed By using Jaccard similarity coefficient (JSC). Characteristic similarity between two services is computed by using weighted sum of Description Similarity and Functionality Similarity. At last services are clustered using Agglomerative Hierarchical Clustering Algorithm.

b) Collaborative Filtering

In this module rating similarity between two services is computed by using Pearson correlation coefficient (PCC).Then neighboring services are selected by using Constraint Formula. In last step all recommended services are ranked in non-ascending order according to their predicted ratings.

III. CLUSTERING

Clustering is a major task in data analysis and data mining applications. It is the method of assigning a objects so that objects in the identical group are more related to each other than to those in other groups. Cluster is an ordered list of data which have the familiar characteristics. Cluster analysis can be done by finding similarities between data according to the characteristics found in the data and grouping similar data objects into clusters. Clustering is an unsupervised learning process. No super-vision means that there is no human expert who has assigned documents to classes. In clustering, it is the distribution and makeup of the data that will determine cluster membership. A good clustering method will produce high superiority clusters with high intra-class similarity and low inter-class similarity. The superiority of a clustering result depends on the similarity measure used by the method and its implementation. The superiority of a clustering technique is also calculated by its ability to find out some or all of the hidden patterns. Similarity of a cluster can be expressed by the distance function. In data mining, there are some requirements for clustering the data. Clustering based collaborative filtering approach mainly contains two types of clustering algorithms.

a) Partitional Clustering

Partitioning clustering algorithm splits the data points into k partition, where each partition represents a cluster. The partition is done based on certain objective function. The cluster should exhibit two properties, these are (a) each group must contain at least one object(b) each object must belong to exactly one group. Partitioning methods relocate instances by moving them from one cluster to another, starting from an initial partitioning. Such methods typically require that the number of clusters will be pre-set by the user. Partitional clustering contains algorithms like K means clustering, K medoids clustering. But these Partitional algorithms have some limitations.

b) Hierarchical Clustering

Hierarchical clustering is a technique of clustering which divide the similar dataset by constructing a hierarchy of clusters. This method is based on the connectivity approach based clustering algorithms. It uses the distance matrix criteria for clustering the data. It constructs clusters step by step. A hierarchical method creates a hierarchical decomposition of the given set of data objects. Tree of clusters is called as dendrograms. Every cluster node contains child clusters, sibling clusters partition the points covered by their common parent. Hierarchical clustering is further divided in to two types.

Agglomerative: Agglomerative hierarchical clustering is a bottom-up clustering method. It starts by letting each object form its own cluster and iteratively merges cluster into larger and larger clusters, until all the objects are in a single cluster or certain termination condition is satisfied. The single cluster becomes the hierarchies root. For the merging step, it finds the two clusters that are closest to each other, and combines the two to form one cluster. Clustering based collaborative filtering approach uses agglomerative algorithm for clustering services.

Divisive: It works in a similar way to agglomerative clustering but in the opposite direction. As it uses top down approach, this method starts with a single cluster containing all objects, and then successively splits resulting clusters until only clusters of individual objects remain.

IV. Recommender System

Recommender systems or recommendation systems are a subclass of information filtering system that used to predict the rating or preference that user would give to an item. Recommender systems typically produce a list of recommendations by using one of two ways - through collaborative filtering or content-based filtering.

V. Collaborative Filtering

Collaborative filtering methods are based on collecting and analyzing a large amount of information on users' behaviors, activities or preferences and predicting what users will like based on their similarity to other users. Advantage of the collaborative filtering approach is that it does not rely on machine analyzable content. It is capable of accurately recommending complex items without requiring an understanding of the item itself. Collaborative Filtering assumes that people who agree in past will agree in future too and people will like the similar kinds if items they like in the past. Collaborative filtering contains two types of techniques, User based collaborative filtering and Item based collaborative filtering.

a) *User Based Collaborative Filtering*

User-based collaborative filtering predicts a user's interest in an item which is based on rating information from similar user profiles. User based CF assumes that a good way to find a certain user's interesting item is to find other users who have a similar interest. This type of technique first tries to find the user's neighbors based on user similarities and then combine the neighbor users' rating scores.

b) Item Based Collaborative Filtering

Item based collaborative filtering technique also applies same idea like user based CF but instead of similarity between users it uses similarity between items. The rating of an item by a user can be predicted by averaging the ratings of other similar items rated by user.

VI. Conclusion

A Clustering based Collaborative Filtering (ClubCF) approach is relevant to service recommendation. Before applying CF technique, services are merged into some clusters via an AHC algorithm. Then the rating similarities between services within the same cluster are computed. As the number of services in a cluster is much less than that of in the whole system, ClubCF costs less online computation time.

Acknowledgement

The author wishes to thank Prof. A. C. Lomte, Prof. G. M. Bhandari and Dr. D. M. Yadav for valuable guidance and encouragement.

References

1. M. A. Beyer and D. Laney, "The importance of ?big data?: A definition," Gartner, Tech. Rep., 2012.
2. Z. Zheng, J. Zhu, M. R. Lyu. "Service-generated Big Data and Big Data-as-a-Service: An Overview," in Proc. IEEE Big Data, pp.403-410, October 2013.
3. A. Bellogín, I. Cantador, F. Díez, et al., "An empirical comparison of social, collaborative filtering, and hybrid recommenders," ACM Trans. on Intelligent Systems and Technology, vol. 4, no. 1, pp. 1-37, January 2013.
4. G. Thilagavathi, D. Srivaishnavi, N. Aparna, et al., "A Survey on Efficient Hierarchical Algorithm used in Clustering," International Journal of Engineering, vol. 2, no. 9, September 2013.
5. J. Wu, L. Chen, Y. Feng, et al., "Predicting quality of service for selection by neighborhood-based collaborative filtering," IEEE Trans. on Systems, Man, and Cybernetics: Systems, vol. 43, no. 2, pp. 428-439, March 2013.
6. M. C. Pham, Y. Cao, R. Klamma, et al., "A Clustering Approach for Collaborative Filtering Recommendation Using Social Network Analysis," Journal of Universal Computer Science, vol. 17, no. 4, pp. 583-604, April 2011.
7. R. D. Simon, X. Tengke, and W. Shengrui, "Combining collaborative filtering and clustering for implicit recommender system," in Proc 2013 IEEE 27th Int'l Conf. on. Advanced Information Networking and Applications, pp. 748-755, March 2013.

AUTHOR(S) PROFILE

Ms. Sheetal Thokal, Received the Bachelors of Engineering degree in Computer Engineering from Pune University and pursuing Masters of Engineering in Computer Engineering from Savitribai Phule Pune University. She is now working as a lecturer in Imperial College of Engineering, Wagholi, Pune.

Prof. Vrunda Bhusari, Received the Bachelors of Engineering degree in Computer Engineering from Nagpur University and Masters of Engineering in Computer Engineering from Bharti University. She is now working as Assistant Professor in Bhivarabai Sawant Institute of Technology and Research, Wagholi, Pune.

