



Hybrid Recommendation System with Review Helpfulness Features

Miss. Jadhav Monika¹, Mrs. Kakade Shital P²

Student, Computer Science & Engg, DACOE, Karad, India ¹

Professor, Computer Science & Engg, DACOE, Karad, India ²

Abstract: As there are millions of product available online it is not easy for the customer to find the best products. Recommendation system is an information filtering system. Many Recommendation system are Collaborative Filtering Algorithm However it still has a problem such as a cold-start problem and sparsity problem. In this paper we proposed Hybrid recommendation system with review helpfulness feature. We can construct the hybrid model. In this system we use the three recommendation technique like as an item based, content based and knowledge based. The result of this system provide the accurate and meaningful service recommendation to active user.

Keywords: Collaborative filtering, big data application, cluster.

INTRODUCTION

In hybrid recommendation system for semantic clusters big data is required as a prerequisite condition. It's having lots of data that exceeds processing capacity of regular database it can unstructured manner and difficult to transport anywhere. In big data application has hugely data collection and it beyond ability of commonly used software to capture to manage and process that data. The most challenge for big data application is explore the large volume of data and remove useful information or knowledge for future action. Recommender system are techniques and intelligent application to guide users in decision making process where they want to select some item alternative services or products.

Recommender system has two main challenges for big data application. 1) within acceptable time to make decision 2) from so many services ideal recommendation can be generated. Big data can be categorized in three ways volume, velocity and variety. Volume is category while surveying the clustering in terms of big data volume is the way of big data. The second category is Velocity it refers the speed of data processing. And last category is Variety which refers no of types of data. In Previous system there may be two problems are occurred Sparsity and new Cold-start problem. In Sparsity Problem this is issue take to place. When the user/items matrix extremely sparse i.e. user can rate only a small no of items so accuracy of recommendation system will be decreased. In most of these system the percentage of the system of rating assigned by user is very small compared to percentage of rating in system has to predict. Hence accuracy of recommender has to predict. In cold-start problem the performance of these system suffers new user or item. When new user enter system it is difficult to find similar one because there is no enough information about user or items history in system.

RELATEDWORK

Hybrid recommendation system studied the clustering methods. In the E-commerce recommendation system designed a neural network based clustering collaborative filtering algorithm. The cluster analysis collect user with same characteristics according to web visiting message data. However, it is difficult to say that user preferences relevant to purchasing preferences. To partition movie k-means clustering algorithm are applied. It can requested by user to provide some extra information it require user to incorporate multidimensional clustering into collaborative filtering recommendation model. In the first stage user and item profile was collected in background data. Using semantic cluster AHC also can clusters similar feature of product/item in second stage. An item prediction was made by performing weighted average of deviations from the neighbours at the third stage. Data providing services can represent in the term of relation between input and output and semantic relation between them. Using the semantic cluster merging the similar services into the same cluster. Network clustering technique on logical network to identify neighbourhood & use CF algorithm to generate recommendation social relationship between this users this work can depends high dimensional parameter used. Hierarchical clustering algorithm it require only implicit feedback on past user purchase to discover the relationship. Then AHC algorithm is applied on services within the same cluster. The content based collaborative filtering approaches having same limitations to avoid these limitations hybrid approach has been introduced. In this approach are combine to gain better performance and eliminate some of drawbacks.

EXISTING SYSTEM

In this existing system when user can enter the system collect the data from user that means user profile and



rating or product list then apply the clustering technique then data can be analysis after the data analysis apply the Item based collaborative filtering. After applying collaborative filtering the system can provide the useful recommendation to the user.

4) Item-based:-Item based recommendation system producing prediction to user Item based algorithm recommends a user the item that is similar to users preferred before.

COLLABORATIVE FILTERING TECHNIQUE

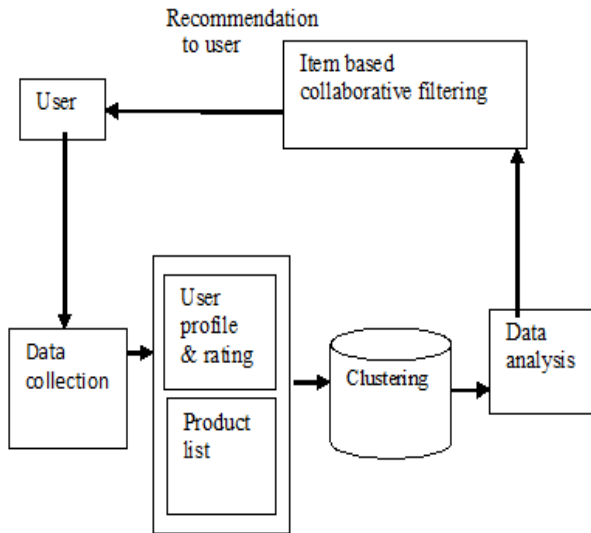
There are two types of collaborative filtering technique as follows:

- 1) Memory based technique.
- 2) Model based technique.

Memory based technique: In memory based technique once a user neighbours is found , the different algorithm can be used to combine neighbours preference to generate recommendation in another way the user were already rated item before relevant role in searching for a neighbour that shares apparition with him. Memory based CF can be achieved in two ways: a) user based technique b) Item based technique.

- a) User based techniques calculated similarly between by comparing same item on their rating.
- b) And item base technique compute predictions using not-similarity between users and similarity between items. It determines how similar retrieved items are to larger items and it selects K most similar items. There corresponding similarities also determined.

Model based technique: This technique can improve performance of collaborative filtering technique. The model building the process by using machine learning data mining technique.



Fig(a). Existing system

RECOMMENDATION TECHNIQUE

There are number of techniques in recommendation system like as collaborative, content –based, demographic, utility based, item -based, knowledge based.

- 1) Collaborative recommendation: - Collaborative recommender system ratings or recommend to objects. User profile in collaborative system consists number of items and ratings, continuously grow up users interacts with system over time. Collaborative filtering is recommended items that similar to users liked.
- 2) Content –based recommendation:- . The basic assumption of content-based system is recommended items that are similar to those the user linked in past. In content-based we need information about items and user profiles. Content based recommendation has user independent, transparent and new item can be easily incorporated.
- 3) Knowledge–based recommendation:- Knowledge based recommendation provide recommendation to the product or services. Knowledge based system needs for knowledge acquisition.

There are three types of knowledge that are:

- a) Catalog Knowledge:-knowledge about objects being recommended.
- b) Functional knowledge:-There need is satisfied to the system which must be able to map between users need and there objects.
- c) User knowledge:-To provide good recommendation system must have knowledge about user.

PROPOSED SYSTEM

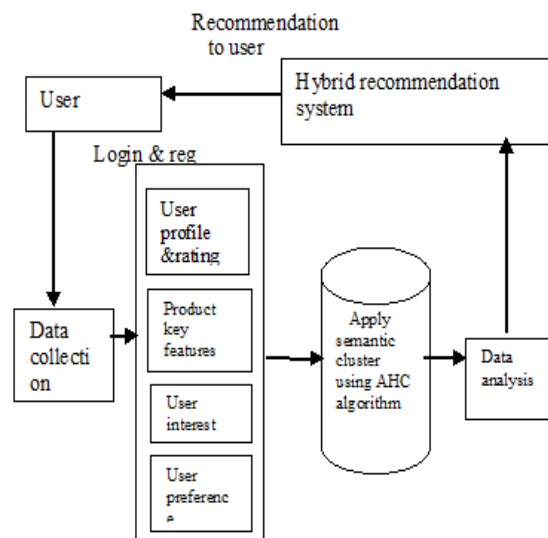


Fig (b).Proposed system

In the proposed system user can first register the data in the system. Then login into the system. After login data will be collect from user that means user profile and



rating, product key feature, user interest, user preferences, after this apply semantic clustering technique by using AHC algorithm [1]. Then data will analysis though the admin. This analysis data goes to hybrid recommendation system. Then this system can be provide the useful recommendation to the user.

AHC Algorithm

- **Agglomerative Hierarchical Clustering (AHC) Algorithm[1].**
- **Input:** A set of services $S = \{s_1, \dots, s_n\}$,
- A characteristic similarity matrix $D = [d_{ij}]_{n \times n}$, the number of required clusters k .
- **Output:** Dendrogram for $k=1$ to S .
- $C_i = \{S_i\}$
- $d_{C_i, C_j} = d_{i,j}$
- for $k = S$ down to K
- Dendrogram $k = \{c_1, \dots, c_k\}$
- $C_1 = \text{Join}(C_l, C_m)$
- for each $C_h \in S$
- if $C_h \neq C_l$ and $C_h \neq C_m$
- $d_{C_l, C_h} = \text{Average}(d_{C_l, C_h}, d_{C_m, C_h})$
- end if
- end for
- $S = S - \{C_m\}$;
- end for

Mathematical model

Let AHC is an Agglomerative Hierarchical Clustering system:

AHC = { I, D, F, R, $D_{sim}(s_t, s_j)$, $F_{sim}(s_t, s_j)$, $C_{sim}(s_t, s_j)$, $R_{sim}(s_t, s_j)$, $ER_{sim}(s_t, s_j)$, O, Φ_p }

- Suppose and are two services such as:
- $S_t = (D_t, F_t, R_t)$ and $S_j = (D_j, F_j, R_j)$
- Where,
- I = Input: in the form of big data.
- $I = \{I_1, I_2, \dots, I_n\} \in \Phi_i$
- D : set of words for describing services ,
- F : set of functionalities ,
- R: set of ratings some users given to item.
- $D_{sim}(s_t, s_j)$: Description similarity,

$$D_{sim}(s_t, s_j) = \frac{|D'_t \cap D'_j|}{|D'_t \cup D'_j|}$$

- $F_{sim}(s_t, s_j)$: Functionality similarity,

$$F_{sim}(s_t, s_j) = \frac{|F_t \cap F_j|}{|F_t \cup F_j|}$$

COMPUTE DESCRIPTION SIMILARITY AND FUNCTIONALITY SIMILARITY

Functionality similarity and Description similarity are both computed by JSC (Jaccard similarity coefficient) which is

a sample set between statically measures of similarity. Jaccard similarity coefficient is defined as intersection of their cardinality divided by union of their cardinality in this description similarity between S_t and S_j is computed by formula

$$D_{sim}(s_t, s_j) = \frac{|D'_t \cap D'_j|}{|D'_t \cup D'_j|}$$

It can be inferred from this formula the larger $|D'_t \cap D'_j|$ s, more similar two services are $|D'_t \cup D'_j|$ is used for scaling factor which ensures that description similarity between 0&1. In the functionalities F_t are gotten from big tableservice. Where the row key = 's_t'

And family of column = "Functionality". In the functionality similarity between s_t and s_j is computed using jaccard similarity coefficient (JSC) as follows

$$F_{sim}(s_t, s_j) = \frac{|F_t \cap F_j|}{|F_t \cup F_j|}$$

CLUSTER SERVICES

In our approach clustering is a critical step these methods partitions a set of object into cluster. Such that same clusters objects are more similar to each other than in different cluster object according to some define criteria. Clustering algorithm can be hierarchical or partitional. There are several limitations for some standard partitional approaches.

1. Results depend on the choice of a clusters K and correct value of K is initially unknown.
2. During the execution of a K means algorithm cluster size is not monitored. Premature termination of an algorithm is caused when some cluster may be come empty ("Collapse").
3. Methods of a hierarchical clustering can be further classified into agglomerative or divisive , depending on the clustering hierarchy is formed in bottom-up or top-down fashion.

CONCLUSION AND FUTURE WORK

In this paper we present a hybrid recommendation techniques for big data application to provide the recommendation services to user before applying collaborative filtering technique, using AHC algorithm the services are merged into same cluster then services within the same cluster. In the whole system the number of services in a cluster is much less. Moreover, as the rating of services in the same cluster are more related with each other. The Rating of services in the same cluster will be more accurate than the rating of all similar or dissimilar services in all cluster.



REFERENCES

- [1] Base paper: Wanchun Dou*, Jianxun Liu, "ClubCF: A Clustering-based Collaborative Filtering Approach for Big Data Application," IEEE, 2014.
- [2] X.Wu, X. Zhu, G. Q.Wu, and W. Ding, "Data mining with big data," IEEE Trans. Knowl. Data Eng., vol. 26, no. 1, pp. 97-107, Jan. 2014.
- [3] T. Niknam, E. Taherian Fard, N. Pourjafarian, and A.Rousta, "An efficient algorithm based on modified imperialist competitive algorithm and K-means for data clustering" Eng. Appl. Artif.Intell., vol. 24, no. 2, pp. 306-317, Mar. 2011.
- [4] J. Michael, Pazzani, Daniel Billsus, "Content-Based Recommendation Systems", Springer-Verlag Berlin Heidelberg, 2007.
- [5] Boucher-Rayan, P. D., Bridge, "Collaborative Recommending using Formal Concept Analysis", Knowledge-based Systems, pp. 309-315, 2006.
- [6] Gunawardana, Christopher Meek, "A Unified Approach to Building Hybrid Recommendation Systems".
- [7] W. Qian, Y. Xianhu, S. Min, "Collaborative Filtering Recommendation Algorithms based on Hybrid User Model", International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 2010.
- [8] U. Farman, S. Ghulam, C. L. Sung, K. P. Yun, D. M. Kyeong, T. K. Jin, "Hybrid recommender System with Temporal information", ICOIN, pp. 421-425, 2012.