International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 2
February 2018

Movie Recommendation System using Enhanced Collaborative Filtering by Fusion Biclusters

***G. Hemalatha**
M.Tech(DCS)
Dept of Computer Science
Pondicherry Engineering College
Puducherry.

****Dr. K. Vivekanandan**
Professor
Dept of Computer Science
Pondicherry Engineering College
Puducherry

Abstract

A movie recommendation is essential in our life since it helps in providing better entertainment. Such a system can suggest good movies to users based on their interest. Although, a set of movie recommendation systems already exist, most of these either cannot predict a movie to the existing users or to a new user efficiently. The proposal is to find movie recommendation system that has the capability to recommend movies even with maximum sparcity. Nearest-neighbor Collaborative Filtering is based either on similarities between users or between items, to form a neighborhood of users or items, respectively. This recommendation system forms biclusters.Then similarity between active user and nearest bicluster is obtained. Later similar users or items and then makes recommendations to individual users. Finally movies are predicted for an active user using item-based CF and user-based CF.

Index terms : Recommender System, Collaborative Filtering, Clustering, Biclusters.

1. Introduction

A movie **recommender system** or a **recommendation system** is a information filtering system that predicts the "rating" or "preference". Recommender systems have become increasingly popular in recent years, and are used in many areas including movies, music, news, books, and social media in general. Recommender systems handle the problem of information overload that users normally encounter by providing them with personalized, exclusive content and service recommendations[01]. Recently, various approaches for building recommendation systems have been developed, which can utilize either collaborative filtering, content-based filtering or hybrid filtering Two main approaches are used for recommender systems. One is content-based filtering, where we try to profile the users interests using information collected, and recommend items based on that profile. The other is collaborative filtering, in which two users share the same interest in the past, e.g if they liked the same book or the same movie, they are likely to have similar tastes in the future. The CF technique can be divided into user-based and item-based CF approaches. In the user-based CF approach, a user will receive recommendations of items liked by similar users[02]. In the item-based CF approach, a user will receive recommendations of items that are similar to those they have loved in the past.

The remaining portion of this paper is structured as follows. Section 2 describes the related work and biclustering techniques. Section 3 describes the limitations of the existing system. Section 4 describes the objectives and features of the proposed model. Section 5 describes the evaluation metrics. Section 6 describes the conclusion.

2. Related Work

In the data mining field , the large volume of complex data are available and many researchs have been carried out on recommendation data using various techniques. Kant, Surya, and Tripti Mahara[01] proposed a biclustering technique to fuse item based and user based bicluster. Zahra, Sobia, Mustansar Ali Ghazanfar, Asra Khalid, Muhammad Awais Azam, Usman Naeem, and Adam Prugel-Bennett[04] A method for finding

initial centroids is proposed by FangYuan et al. In this method, centroids are produced systematically, which are consistent with distribution of data. Though no improvement in time complexity is proposed but it produces better quality clusters[04]. Fang Yuan Fang Yuan et al. provide comparison of 14 k-means cluster initialization methods, which come under two main categories—synthetic starting points and actual sample starting points. Faris Alqadah·Chandan K. Reddy·Junling Hu·Hatim F. Alqadah[06] The set of biclusters in matrix K are ordered by the hierarchical order; under this ordering the set of biclusters form a complete lattice[06]. Using this formulation identifying neighborhoods of closely related biclusters is a well-defined task and several efficient algorithms exist to identify these neighborhoods. Using several important properties from the field of Formal Concept Analysis, building user-specific biclusters that are "more personalized" to the users of interest[06]. By exploring the hierarchical relationship between one bicluster with other biclusters, we construct bicluster similarity to capture local proximity between a recommended item and a user.

User based Collaborative Filtering using fuzzy C-means Hamidreza Koohi, Kourosh Kiani. A fuzzy C-means approach has been proposed for user-based Collaborative Filtering and its performance against different clustering approaches has been assessed. The MovieLens dataset is used to compare different clustering algorithms[07]. They are evaluated in terms of recommendation accuracy, precision and recall. The empirical results indicate that a combination of Center of Gravity defuzzified Fuzzy Clustering and Pearson correlation coefficient can yield better recommendation results, compared to other techniques.[07]. Personalized electronic movie recommendation system based on support vector machine and in[08] The proposed PRS not only considers the movie's content information but also integrates the users'demographic and behavioral information to better capture the users'interests and preferences. The efficiency of the proposed method is verified by a series of experiments based on the MovieLens benchmark dataset.proved particle swarm optimization[08]by IibinWang. FengjiLuo YingQian GianlucaRanzi. Recommender systems for large-scale e-commerce: Scalable neighborhood formation using clustering[09] Sarwar, Badrul M., George Karypis, Joseph Konstan, and John Riedl A comprehensive algorithmic framework supported by various techniques of information retrieval. A well-rounded methodology that explores concepts of data ,information, knowledge ,and relations between them to support a formation of a suitable recommendation[09]. In particular, the developed system helps collecting data characterizing potential reviewers ,retrieving information from relational and unstructured data ,and formulating a set of recommendations.

3. Limitations of the Existing System

The existing approach called NBCF(Nearest Bicluster CF),Xmotif biclustering algorithm is applied to group the similar users/items to create overlapping clusters.Next nearest bicluster is computed by finding similarity measures, and later prediction.

The limitations in existing system are :

- X-motif takes random users and therefore all users are not taken.
- Does not perform well for overlapping bi-clusters and hence prediction accuracy is low
- Does not works well for additive bicluster

4. Objectives and features of the Proposed Model

The primary objective is to suggest a good movie recommendation through data clustering for the customers. Good recommendation systems should be able to collect information about user's choices, and can use this information to improve their suggestions in the future. The objective of my work is to produce good recommendation even in sparse environment. Hence, the following:

- Propose to fuse the item-based CF and user-based CF.
- To find a new Biclustering technique which works well in sparse environment

4.1 Novel Summation Bicluster

Novel summation bicluster is a technique which includes all the users who have given ratings thereby increasing the prediction accuracy and avoiding overlap clusters .The novel bicluster technique has strong

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 2
February 2018

partial similarity with an active user. Since this technique has partial similarity and includes all the users prediction is accurate. The item rating may overlap in clusters for each individual users .First one user is selected and all items with rating for tat user is summed up and the same item for all other user is summed up to form a bicluster .The procedure is repeated for all other users except who are included in previous formed cluster.

ARCHITECTURE OF PROPOSED SYSTEM

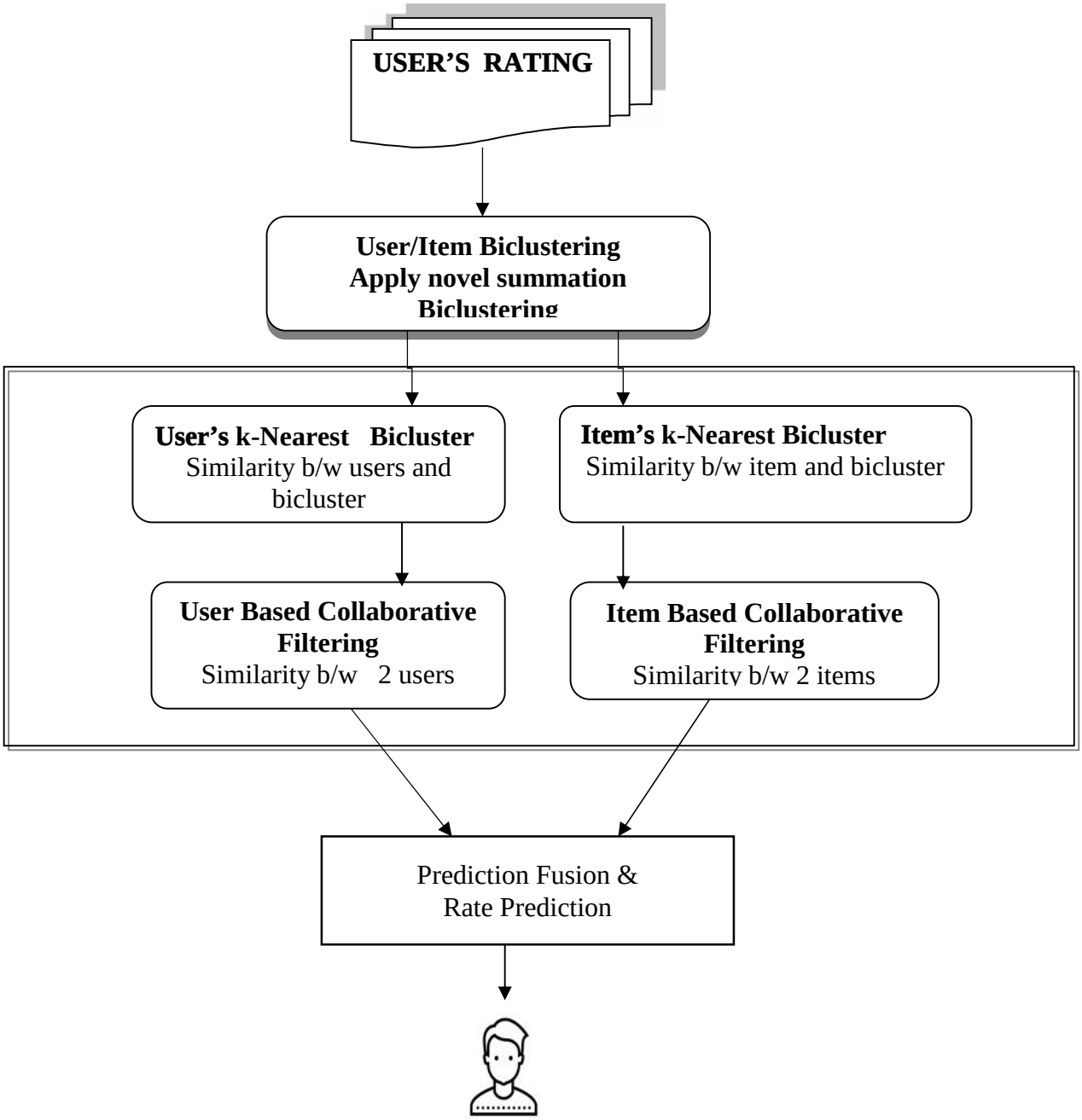

Figure 1: Architecture of Proposed System

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 5, Issue 2
February 2018

5. Evaluation Metrics

To measure the rating prediction accuracy of the recommendation algorithms, two standard metrics namely **Root Mean Squared Error (RMSE)** and **Mean Absolute Error (MAE)** are frequently used. The MAE, is a measure of the average absolute deviation of recommendations from their actual rating to the predicted rating.

MAE :

$$\mathbf{MAE} = \frac{1}{|I|}\sum_{i=1}^{|I|} |p \qquad - r|$$

RMSE:

$$\mathbf{RMSE} = \sqrt{\frac{1}{|I|}\sum_{i=1}^{|I|}(p \qquad - r)}$$

Where, |I| - total number of test ratings, Pred - predicted value, r- actual value.

Conclusion

CF remains one of the most popular and widely accepted methods that can handle the information overload problem effectively. It aims at suggesting suitable items for a user based on rating information collected from other similar users. Although, they are very successful and popular in many areas, they often confront the sparsity problem. The proposal is to merge the item-based and user-based CF in a weighted sum approach .To find the nearest bicluster of active user, a new similarity approach is proposed. For future work, it would be interesting to test the system on other datasets, with a new similarity measure or something else to obtains a bicluster that has strong partial similarity with an active user's preferences.

References

[01] Kant, Surya, and Tripti Mahara. "Nearest biclusters collaborative filtering framework with fusion." *Journal of Computational Science* (2017).

[02] Isinkaye, F. O., Y. O. Folajimi, and B. A. Ojokoh. "Recommendation systems: Principles, methods and evaluation." *Egyptian Informatics Journal* 16, no. 3 (2015): 261-273.

[03] Bobadilla, Jesús, Fernando Ortega, Antonio Hernando, and Abraham Gutiérrez. "Recommender systems survey." *Knowledge-based systems* 46 (2013): 109-132.

[04] Zahra, Sobia, Mustansar Ali Ghazanfar, Asra Khalid, Muhammad Awais Azam, Usman Naeem, and Adam Prugel-Bennett. "Novel centroid selection approaches for K-Means-clustering based recommender systems." *Information sciences* 320 (2015): 156-189.

[05] Sarwar, Badrul, George Karypis, Joseph Konstan, and John Riedl. "Item-based collaborative filtering recommendation algorithms." In *Proceedings of the 10th international conference on World Wide Web*, pp. 285-295. ACM, 2001.

[06] Alqadah, Faris, Chandan K. Reddy, Junling Hu, and Hatim F. Alqadah. "Biclustering neighborhood-based collaborative filtering method for top-n recommender systems." *Knowledge and Information Systems* 44, no. 2 (2015): 475-491.

[07] Hamidreza Koohi, Kourosh Kiani. " User based Collaborative Filtering using fuzzy C-means." *Computational Intelligence* 32, no. 4 (2016): 561-586.

[08] Wang, Xibin, Fengji Luo, Ying Qian, and Gianluca Ranzi. "A personalized electronic movie recommendation system based on support vector machine and improved particle swarm optimization." *PloS one* 11, no. 11 (2016): e0165868.

[09] Sarwar, Badrul M., George Karypis, Joseph Konstan, and John Riedl. "Recommender systems for large-scale e-commerce: Scalable neighborhood formation using clustering." In *Proceedings of the fifth international conference on computer and information technology*, vol. 1. 2002.

[10] Kim, Kyoung-jae, and Hyunchul Ahn. "A recommender system using GA K-means clustering in an online shopping market." *Expert systems with applications* 34, no. 2 (2008): 1200-1209.