An Adaptive Clustering and Incremental Learning Solution to Cold Start Problem in Recommender System using K-Means Method

K.Yogeswara Rao[1], B.Dhanushya[2], G.Praveen Kumar[3]
Aditya Institute of Technology and Management

Abstract- Recommender systems are the systems that recommend items to the users based on their tastes and interest. Many e-commerce websites use recommender systems to recommend items to users. With the expansion of data on the web and a variety of products available in the E-commerce sector, the recommender systems come to the aid of the users. When a user or item is new, the system may fail to give best possible results because of insufficient information availability on this item. When a new user enter the system and does not find recommendations that are not convincing him, the user may not reuse the system and market value of the system drops down. This is called Cold Start Problem. Various solutions to this 'cold start problem' have been proposed in the literature. However, many real-life E-commerce applications suffer from an aggravated, recurring version of cold-start even for known users or items, since many users visit the website rarely, change their interests over time.

The cold start problem is one of the major drawbacks for the recommender system and can prove costly for its success if proper steps are not taken. We solve this by proposing variations in adaptive clustering methods that will be used in the pre-processing stage to generate data only for lowly ranked items so that they can compete with the highly ranked items. So by using adaptive clustering technique we separate highly rated and lowly rated datasets, in addition to this we extract the medium rated data items from previous highly rated and lowly rated data items by using k-means algorithm. So that we can avoid the cold start problem to some more extent.

Keywords- Recomender Systems;coldstart problem;highly ranked items ; medium ranked items;lowly ranked items.

I. INTRODUCTION

A. Recommender systems: Recommender systems [1, 3, 4, 6, 10, and 13] are the systems that recommend items to the users based on their tastes and interest. With the expansion of data on the web and a variety of products available in the E-Commerce sector, the recommender systems come to the aid the users. Some of the websites that use recommender systems are Amazon.com and Ebay.com for product sales, Google and Yahoo! for advertisements, IMDB.com and MovieLens.com for movie recommendations and so on.

Recommender systems have become an important research area since the appearance of collaborative filtering in the mid1990s. The interest in this area still remains high because it constitutes a problem rich research area and because of the plethora of practical applications that help the users to deal with vast information and provide personalized recommendations.

The Recommender systems have been aggressively incorporated as a core part to most online shops. They help users who are going through numerous items like movies or music presented in online shops by quickly capturing each user's preference on items and suggesting a set of personalized items that he/she is likely to prefer. The recommendations are predicted based on the features of the neighborhood of the active target user. To determine the similarity between the items/users, there are many similarity functions exists and Pearson Correlation Coefficient is most commonly used similarity metric.

***B.Cold start problem*:** The cold start problem [2,3,7,8, and 16] is one of the major drawbacks for the recommender system and can prove costly for its success if proper steps are not taken. When a new user enters the system and does not find recommendations that are not convincing him, the user may not reuse the system and market value of the system drops down. Some alternatives need to be framed out to handle the new customers and keep the performance levels high. Similarly, for a new item introduced in the system, if it does not get make it to the recommendations of users the suppliers will lose trust on the system and will not incorporate it in their businesses. For a new item or items with less data, they will never fall in the top-K recommendations to the user even if they are desired in the recommendations and hence their rating count from the user would never improve. This is called cold start problem.

The cold start problem for items is popularly known as the long tail problem as the items in the item list can be classified into two portions called head and tail. The head bears the items with large number of ratings from the user while the tail bears the remaining items with insufficient ratings. By the study of real life datasets it has been learnt that the number of items in the tail are far greater than the items in the head and hence the name long tail problem is popular.

Collaborative filtering: Collaborative filtering [1, 2, 3, 6, 8, 11 and 14] exploits the ratings on items given by other users whose preferences are similar to a target user, is one of the most successful and widely-used methods to learn the connection between users and items. Collaborative Filtering systems are information retrieval systems that operate under the assumption a user will like the same data items that other users have liked in the past. These systems are particularly popular and have been applied in many online shopping websites. Collaborative filtering algorithms mainly aggregate feedback for items from different users and use the similarities between items and items or between users and users to provide recommendation to a target user.

Collaborative filtering can be categorized into two techniques: neighborhood models and latent factor models. Neighborhood models find k most similar users i.e., nearest neighbors, to a target user and make a prediction of a target user on an item based on the rating patterns of those neighbors on the item. Neighborhood models are sub-categorized into user-based models and item-based models. On the other hand, latent factor models learn hidden from observed ratings by using factorization techniques. Collaborative Filtering is one of the methods widely used in recommender systems using two different techniques, memory-based and model-based. The memory-based depends on the entire rating which exists in the user-item matrix for forming neighbors of the active user to generate recommendation tailored to his/her preferences. In contrast, the model-based methods learn the models of recommendations from the entire ratings to generate the recommendation for the target user.

II. RELATED WORK

A. Problem Definition:

To solve the cold start problem by proposing variations in adaptive clustering methods that will be used in the pre-processing stage to generate data only for lowly ranked items so that they can compete with the highly ranked items

. • To develop a predictor based on incremental learning algorithm based on LEARN++ that meet the needs of the adaptively clustered datasets.

• In addition to this we extract medium rated data items from above data items, so that we can reduce cold start problem to more extent.

B. Objectives of the Problem :

• Objective 1: To separate medium rated items from highly rated and lowly rated items by using k-means algorithm.

• Objective 2: To reduce cold start problem in medium rated items also so that they can also be competed with highly rated items.

C.Algorithm:

```
Algorithm:
            Import the Electronic items dataset from kaggle which is
Amazon consumer review dataset.
K_means(D,k)
            where D is the dataset we are taking and k indicates no. of clust
we want which is 2 since we want to get 2 clusters one is lowly rated items and ot
is highly rated items
Step1:- From data take two random mean
        values m1 and m2.
Step2:- for i=1 to n do
        for j=1 to n do
         for each Data item Di belongs to D do
if(Di~m1)<(Di~m2)
        then
         Ki={Di} Associate Di into the Ki cluster
        Else
Lj={Di} Associate Di into the Lj cluster
        end for
Step 3:- m1=? Ki/n(Ki)     m2=? Lj/n(Lj)
Step 4:- Repeat step 2 and 3
         until Ki=Ki-1 &&Lj=Lj-1
          end for
          end for
Step 5:-We got 2 clusters Ki which is lowly rated
         cluster and Lj which is higly rated cluster
Step 6:- Now evaluate K_means(ki,2),
K_means(Lj,2)
Step 7:- From above evaluation we get 4 cluster
Kii, Kij, Kji, Kjj
        we merge Kij and Kji we get
         Medium rated dataset=[kij+kji]
```

III. EXPERIMNTAL EVALUATION

A.Dataset

Electronic items data set from kaggle Website

https://www.kaggle.com/datafiniti/consumer-reviews-of-amazon-products

B.*Result Analysis*

In order to solve the cold start problem in items, various adaptive clustering improvements were proposed and tested for performance. Two improvements are suggested to improve the adaptive clustering process i.e. Adaptive clustering with Pearson's similarity and Adaptive clustering with Slope One predictor. Each of these methods tries to improve the existing clustering process and reduce the error rates in the lowly ranked items.

An incremental learning algorithm also has been proposed to suit the needs of adaptive clustering. Using this algorithm, the weight age of adaptively clustered data can be reduced over time and the system can make a smooth transition from pre-processed data towards the true data that is provided by the customers. A study of the improvement of error performance of the proposed methods has been done and tabulated as shown below.

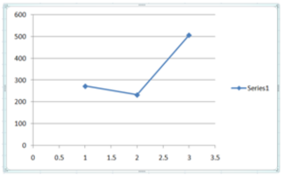
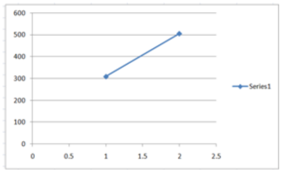
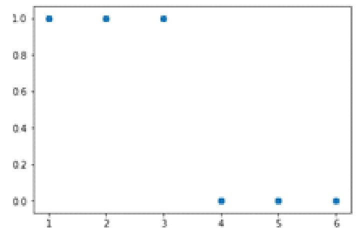
Adaptive clustering algorithm for lowly rated and highly rated items	K-means clustering algorithm for medium rated items
Lowly rated items: 54.07% Highly rated items: 46.12%	Medium rated items: 61.43%
	

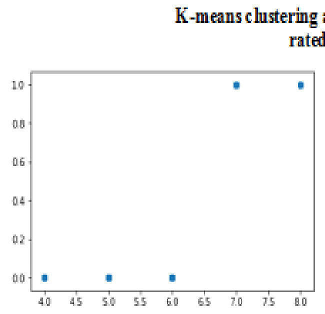
Fig- a Experimental Comparison

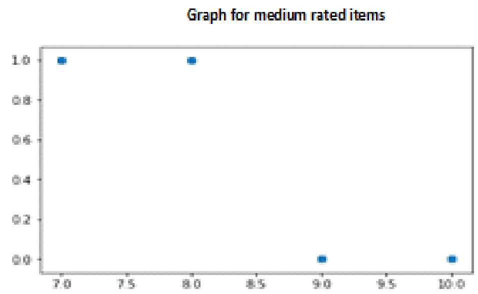
A. \ Graphical Analysis:

Adaptive clustering algorithm for lowly rated and highly rated items



Graph for low rated items





Graph for highly rated items

Fig-b. Graphical Comparison

IV. CONCLUSION

In order to solve the cold start problem, improvements have been suggested for the adaptive clustering process that shows considerable improvement in the error rates of the lowly ranked items. An incremental learning predicator is also proposed that learns from time to time and incorporates the learning of new information. The study of adaptive clustering is limited to the selection of α.So that we have generated medium rating items from lowly rated and highly rated items by using K-means clustering algorithm. So that we can generate data only for medium rated items so that we can avoid cold start problem to some more extent. The future work could involve work on dynamic change of α for appropriate clustering of data arriving from data streams.

V. REFERENCES

[1]. Ferran Torrent-Fontbona, Joaquim Massana, Beatriz Lopez (2018, November). 'Case-base maintenance of a personalized and adaptive CBR bolus insulin recommender system for type 1 diabetes' University of Girona, Campus Montilivi, building EPS4, Girona, Catalonia.

[2]. Cold Start Mechanism To Prevent Compromise Of Automatic Anomaly Detection Systems, by Fabien Flacher et al., October 8, 2015.

[3]. On identifying K-nearest neighbors in neighborhood models for efficient and effective collaborative filtering. December 4, 2016.

[4]. Recommender system based on pairwise association rules Timur Osadchiy, Open Lab, School of Computing, Newcastle University, Newcastle upon Tyne, United Kingdom, 4th April, 2018.

[5]. An Adaptive Large Neighbourhood Search Algorithm for the Orienteering Problem by Alberto Santini, UniversitatPompeuFabra and Barcelona GSE, 08005 Barcelona, Spain, 12 th January 2019.

[6]. A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques by MehrabakshNilashi, Othman Ibrahim, Faculty of Computing, UniversitiTeknologi Malaysia, Johor, Malaysia, 22 July, 2017.

[7]. Simultaneous co-clustering and learning to address the cold start problem in recommender systems Andre Luiz Vizine Pereira, University of Sao Paulo, Brazil, 6 Septmber, 2014.

[8]. Adaptive Collabortive Filtering Based on Scalable Clustering for Big Recommender Systems by O-Joun Lee, Min-Sung Hong, and JasosnJ.Jung, School of Computer Engineering Chung-Ang University, Korea 8, August 2015

[9]. Incremental maximum margin clustering by V.VijayaSaradhi. P. Charly Abraham, 27th January 2014.

[10]. Adaptive KNN based Recommender System through Mining of User Preferences by V.Subramaniyaswamy, March, 2017.

[11]. Dynamic Collaborative Filtering System Via a Weighted Clustering approach by Agiles Salah, Universite Paris Descartes, France 15th May 2015.

[12]. Collaborative Filtering-Based Recommender System: Approaches and Research Challenges, by Ritu Sharma, Department of Computer Science & Engineering, Malaviya National Institute of Technology, Jaipur, India 13th September, 2017.

[13]. A population-based incremental learning approach with artificial immune system for network intrusion detection by Meing-Hui Chen, School of Software, Nanchang, Chibna, 2011.

[14]. Adaptive Collaborative Filtering With Extended Kalman Filters and Multi-Armed Bandits by the French Government, 28th September, 2017.

[15]. An adapted incremental graded multi-label classification model for recommendation systems by Khalil Laghman, Gerrmany, 7th June, 2017.

[16]. An Efficient k-Means Clustering Algorithm: Analysis and implementation by Tapas Kanungo, Senior Member, IEEE 7th July, 2002.

K Yogeswara Rao received his M.tech From Andhra University, Visakhapatnam AP India , and PhD in Jan 2017 in Computer Science and Engineering and area of Specialization is Text Mining. He is working as a Associate Professor in Aditya Institute of Technology and Management, Tekkali-532201, AP,India. He is having 19+ Years of experience in teaching. He published good number of papers in International Journals with good impact factor and Scopus Indexed. He is a Member of CSI.

I am B.Dhanushya. I am pursuing B.Tech at Aditya Institute of Technology and Management, Tekkali I am good at C, Java, Python and R-Programming.. I got placed as a Project Engineer in Wipro Technologies in our campus recruitment.

I'm G.PRAVEEN KUMAR from CSE Department in AITAM college,Tekkali.I have completed a training program on "Android 7.0 Application Development with Material Design using Android Studio". I have done my internship at AQURA INFOTECH on ANDROID in Vishakhapatnam. I successfully completed courses on "C,C++,Core JAVA, Advanced JAVA & Oracle, HTML, CSS, PHP, Boot Strap".