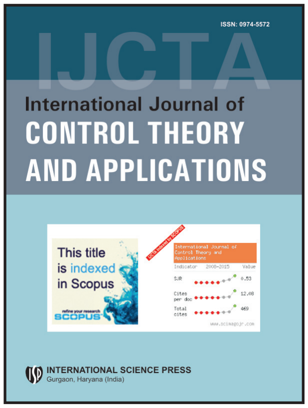

International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 10 • Number 12 • 2017

Mining Big Data: Towards a Machine Learning Framework Based on Collaborative Filtering

T. Venkatesan[a], T. Ramkumar[b] and K. Saravanan[c]

[a]*Research Scholar, Department of Computer Science, PRIST University, Thanjavur*
[b]*Assoc. Professor, School of Information Technology & Engineering, VIT University, Vellore*
[c]*Dean, Faculty of Computer Science, PRIST University, Thanjavur*

Abstract: Modern computing and communication devices open the venue for generating flow of data in many forms which are characterized in-terms of three main features includes *volume*, *velocity*, and *variety*, deemed as 'Big Data'. Though quantum of data generated from various sources in heterogeneous format, it is mandatory to analyze such huge data for reaping meaningful patterns of interest towards decision making process since an effective central and branch level corporate decision are always based on the data coming from diversified data sources. In this paper, we have focused the problem of mining patterns of interest from the big data and adopt the machine learning framework for recommendation of items to the user on the basis of personalized preferences. Now a day, recommender system has been widely accepted as a strategy by organization for predicting their consuming behavior and employed as a tool for effective customer relationship management. We have also studied various research works in the recommender system and implement the item based recommender using a bench marked dataset namely, *movie-lens* dataset through the experimental set-up conducted using Hadoop and Mahout Machine learning library.

Keywords: Big Data, Collaborative Filtering, Recommender System, Machine Learning, Mahout, Big Data Mining.

1. INTRODUCTION

Big Data refers the strong growth of heterogeneous information flows due to the increase in usage of new technologies and computing methodologies. With the abundant growth of data sources like social networks, mobile, and IoT devices, massive amount of information has been accumulated from those diversified data sources. Hence mining pattern of interest from multiple heterogeneous data sources significantly creates promising research avenues among the data science community [1, 2, 6]. The implication and outcome of the big data has been successfully recognized in various domains such as healthcare and bio-medical (eg. Towards enhancing the efficiency of treatment methods and personalized patient health care system), Sales and marketing (eg. Targeting the sales campaign for the proposed promotion), Transportation (eg. Identifying optimal routes with reduced costs towards speedy service), Finance (eg. Idenfying low credit risk groups for offering loans), Information technology (eg. Identification of social community forums with similar likes) and E-governance (eg. By creating societal data repositories for customized user interaction queries).

When such kind of big data is being emerged, it is appropriate to mine those for identifying the various kinds of patterns in the respective application domain [3, 4]. In such circumstances, the retrieval of meaningful, potentially useful patterns of interest on the basis of user preferences plays a vital role. Though information retrieval systems such as search engines prioritize the retrieved information, they are unable to provide pattern of information on the basis of user preferences or personalized interest [5]. To enforce personalization of users, we have advocated a machine learning approach for mining big data on the basis of the recommendation system framework. In the context of big data, recommendation system is very useful for both kinds of internet users such as service providers and service consumers. Due to the changing nature of service consumers, offering various kinds of services on the basis of user preferences/personalization is the central requirement of any consumer centric business or applications.

Recommender system can be considered as a decision supportive framework that facilitates strategies and conventions to the service users for obtaining optimal recommendations towards availing the resources or services on the basis of the personalization or preferences of the user of concerned. With this nature, recommender system plays a significant role in Customer Relationship Management (CRM) based application, where with the inclusion of recommender, the cost and time for identifying sustainable consumers and services have been significantly reduced. The general recommender framework falls under the following categories namely, (i) Content based filtering – where, the recommendation is being performed by matching or correlating the content of the service or item against the characteristics or features of the consumers or users. (ii) Collaborative filtering - where, the services or items have been recommended to the consumers or users on the basis of the taste of the similar kinds of users. (iii) Hybrid filtering – where, recommendation has been performed by combining both collaborative and item based filtering approaches. This approach creates an unified recommendation system by applying collaborative filtering techniques on content based approaches and it applies content based filtering techniques on collaborative approaches. Hence an effective recommendation framework has been resulted.

The broad taxonomy recommendation system along with various filtering techniques is shown in Figure 1. At layer 1, the system has been classified on the basis of three filtering techniques namely; content based filtering, collaborative filtering and hybrid filtering techniques. Further specialization of each filtering techniques are devised in layer 2. The content based filtering technique attempts the recommendation of items by correlating the item contents against the user profile. The profile of the user is being compared with the features of the content of the item where the user has made evaluation already. Here, the profile of the targeted user alone is required and it doesn’t need profile of similar users for the task of recommendation. Techniques based on vector space model (Term Frequency/Invert Document Frequency), probabilistic models (Naive Bayes Classifier), machine learning models (Decision Tree) have been widely used in this approach. Recommending web pages, news items are the appropriate examples for the content-based filtering.

As we stated earlier, the collaborative filtering approach recommends the items by comparing the taste of similar user. Here the item/user similarity matrix has been constructed for obtaining the items preferred by the active user (user, who needs recommendation). This approach has been further divided in to memory based techniques and model based techniques as shown in level 3. Memory based techniques use the entire or sample of the user/item matrix to generate the recommendation. It has been implemented in-terms of user based and item based filtering techniques. In user based technique as shown in Figure 2, user/item matrix has been used for identifying similarity between the users by comparing their rating on the same items and prediction of rating for the active user has been made.

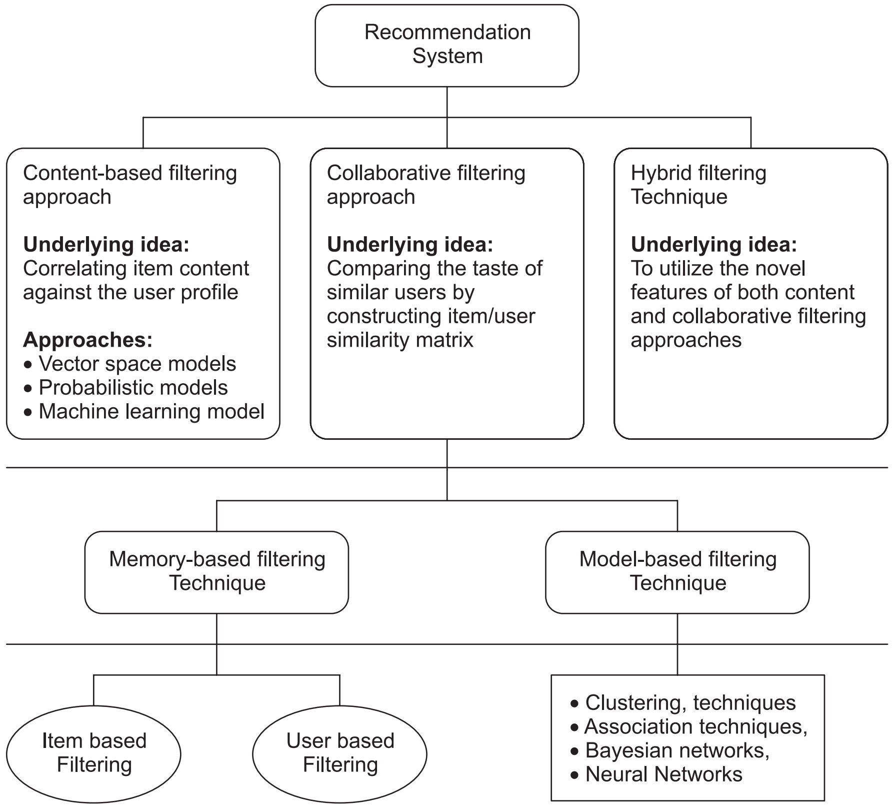

Figure 1: Recommendation techniques

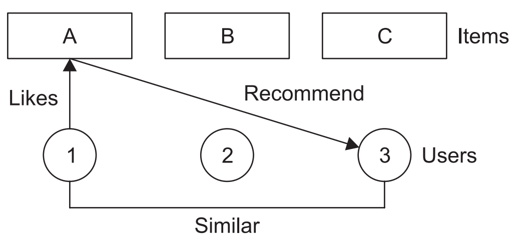

Figure 2: User-based Collaborative Filtering

In item based recommendation techniques, the user/item matrix has been used to extract the similarities among the items rated by the active user as shown in Figure 3.

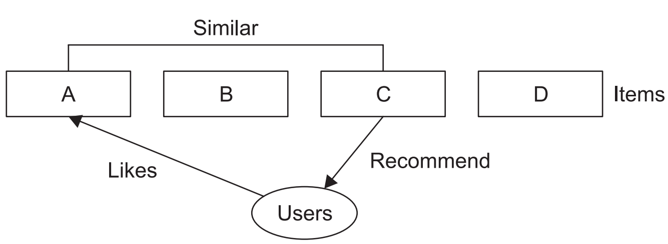

Figure 3: Item-based Collaborative Filtering

The above two memory based recommendations systems are not suitable for the task of recommendation when the targeted data set is very large. Hence, the model based techniques has been widely used for applications which deals with large datasets. In model based techniques, a recommendation model has been trained on the basis of previous rating given by the similar kinds of users. Descriptive data mining techniques such as clustering, association rule mining, predictive data mining techniques such as classification, regressions have been widely applied in the model construction process.

In this paper, we have established a big data analytical framework using machine learning library called 'Mahout', for predicting recommendations of items by the user. The structure of the paper has been conceived as follows; In section 2, various state-of-the art research attempts on recommendation systems have been highlighted along with their research outcomes. Besides, various successful online recommendation systems have also been mentioned. In section 3, a big data framework has been established on the basis of Hadoop platform and mahout machine learning library for providing a recommendation framework based on the benchmark data set called *Movie Lens*, and section 4 concludes the paper.

2. RESEARCH OUTCOMES ON RECOMMENDATION SYSTEMS

This section covers the features of various recommendation systems that have shown its success for providing online recommendation services. Also it accounts the various research works in the broad area. Tapestry[7] is the one of the earliest implementation of collaborative filtering based recommendation system which outcomes the opinion of the closed newsgroup community as recommendation. It suffers the significant drawback such as inclusion of complicated queries for performing recommendation. A News based recommendation architecture, called Group Lens[8] that implements collaborative filtering methods has been developed for the provision of required news from the large amount of news database based on their personalization. Ringo[9] is an online music recommender system that works on the basis of collaborative filtering approaches. Amazon[10] has been witnessed as pioneer in the usage of recommendation framework for recommending items to the user based on the browsing history of similar users. The content based online book recommendation system, LIBRA[11] is used to recommend the suitable titles to the user on the basis of trained content studied from the various kinds of user. News Dude [12]is a personal news system, based on Term Frequency-Inverse Document Frequency (TF-IDF) model used to recommends news stories to users. The most popular citation index recommender, Cite Seer uses various heuristic and machine learning algorithms for performing recommendation of citations on the basis of preferences of researcher**.** A programming tutorial recommendation system called PROTUS has been developed by Boban Vesin et. al., [13]. The proposed system recommends the courses to the students based on the features like age, domain of study, and others.

Konstas et. al., [14] proposed a music recommendation system based on the features like tagging, play hit for recommending the social relation of the user. Cunningham et. al., [15] proposed an hybrid recommendation system which combines content and collaborative filtering approaches. Ghazanfar and Pragel-Bennet[16] proposed a hybrid recommendation approach which uses content based profile of individual users to find similar user for the task of recommendation. In Sarwar et. al., [17] work, an information filtering agent has been introduced in the existing content based filtering. Accordingly, a framework has been proposed for integrating content filtering approaches with collaborative techniques.

Since the amount of big data is being accumulated in a huge manner, a single machine even with substantial amount of memory is also not feasible to process the data. Hence the usage of distributed systems for performing recommendation task over big data has been paid more attention by the research community. Such kind of distributed environment has been offered by the open source distributed framework called Hadoop, most of the recommendation algorithms have been implemented on big data environment with less processing time. Lin et. al., [18] advocated a distributed approach for performing recommendation on big data. Their distributed collaborative filtering recommendation combines K-Means algorithm and slope-one recommendation on hadoop. Shunmei Meng et. al., [19] proposed a scalable machine learning recommendation framework called KASR (A Keyword-Aware Service Recommendation) method for big data based on collaborative filtering algorithm. For finding the similarities, they have used Jacord co-efficient and cosine similarity measures. Zhao and Shang [20] have implemented collaborative filtering using Hadoop distributed framework that offers scalable recommendation system for big data. DePessemier et. al., [21] presented a recommendation framework based on Hadoop for calculating content-based recommendation and pair wise item/user similarities. In their experimental set-up, they have taken the Wikipedia documents and performed various content based suggestions to the user. The above promising research works attempted in the big data context ensures the novelty of various recommendation algorithm for the provision of personalized pattern of interests.

3. IMPLEMENTATION SCENARIO

This section mainly focuses on implementation results of the Item – Based Collaborative Filtering using Pearson correlation coefficient similarity using Mahout Library's on the Hadoop platform. As an open source machine learning library, Mahout offers wide range of applications that are useful for the task of recommendation. It provides collaborative filtering algorithm, data clustering, and data classification. The scalability nature of mahout able to support distributed processing of large data sets across cluster of nodes using Hadoop Distributed File System (HDFS). To calculate similarity among the object, it does various similarity calculations and evaluation methods such as Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). As an open source project, Mahout has an active user community that periodically updates the code.

3.1. Dataset

For the experiment, we have used the *MovieLens100k* dataset (*http://grouplens.org/datasets/movielens/).*The dataset contains 1, 00,000 preferences applied to 1682 movies by 943 users of the online movie recommender service. The data are contained in three files, movies.dat, ratings.dat and tags.dat. Rating data files have at least three columns: the user ID, the item ID, and the rating value. Each user at least rates 20 movies; the range of ratings is from 1(lowest) to 5(highest).

3.2 Similarity Computation

Similarity computation among items or users plays a vital role in memory-based collaborative filtering algorithms. For item-based based algorithms, the user/item matrix has been used to extract the similarities among the items rated by the active user. To find the similarity among items, we have used Pearson correlation coefficient.

$$\mathrm{W}_{a,u} = \frac{\sum_{i=1}^{n}(r_a, i - \bar{r}_a)\cdot(r_u, i - \bar{r}_u)}{\sqrt{\sum_{i=1}^{n}(r_a, i - \bar{r}_a)^2}\cdot\sqrt{\sum_{i=1}^{n}(r_u, i - \bar{r}_u)^2}}$$

where,

a: the active user

u: another of the users of the system

n: the number of items that both the active user and ALL recommender users have rated

r_a: is the average ratings of the active user r_u is the average of another user 'u' 's' ratings

$w_{a,u}$: the degree of correlation between user in 'a' and user 'u'.

Then the prediction for user 'a' on item 'i' denoted by $p_{a,i}$ is calculated as follows:

$$P_{a,i} = \frac{\bar{r}_a + \sum_{u=1}^{m} (r_u, i - \bar{r}_u) \times W_{a,u}}{\sum_{u=1}^{m} \left| W_{a,u} \right|}$$

where, m is the total number of users.

The experimental set up has been created in a Hadoop Version of 1.2.1 with the JDK of 1.6. The mahout library version of 11.00 has been used. The experiments are evaluated for Item-Based Collaborative Filtering that based on Pearson Correlation Coefficient Similarity algorithms. Initially, we have framed the User-item based matrix for performing the item based recommendation for the movie-lens data set. Then Pearson Correlation coefficient technique is being used for calculating the degree of correlation among the rated and non-rated items by a user. The active user 'a' whose recommendation for the non-rated item has been calculated with the help of the rating made by the other item, denoted as 'u'. For an example, in the data set movie lens, user 'u1' has made his recommendation only for the 273 items out of 1682 movie items and the details are shown in Table 1.

Table 1: Actual items recommended by the Active User '1'

User id	*Movie id along with rating*
1	61:4,189:3,33:4,160:4,20:4,202:5,171:5,265:4,155:2,117:3,47:4,222:4,253:5,113:5,227:4,17:3,90:4,64:5,92:3,228:5,266:1,121:4,114:5,132:4,74:1,134:498:4,186:4,221:5,84:4,31:3,70:3,60:5,177:5,27:2,260:1,145:2,174:5,159:3,82:5,56:4,272:3,80:4,229:4,140:1,225:2,235:5,120:1,125:3,215:3,6:5,104:1,49:3,206:4,76:4,72:4,185:4,96:5,213:2,233:2,258:5,81:5,78:1,212:4,143:1,151:4,51:4,175:5,107:4,218:3,209:4,259:1,108:5,262:3,12:5,14:5,97:3,44:5,53:3,163:4,210:4,184:4,157:4,201:3,150:5,183:5,248:4,208:5,128:4,242:5,148:2,112:1,193:4,264:2,219:1,232:3,236:4,252:2,200:3,180:3,250:4,85:3,91:5,10:3,254:1,129:5,241:4,130:3,255:2,103:1,118:3,54:3,267:4,24:3,86:5,196:539:4,164:3,230:4,36:2,23:4,224:5,73:3,67:3,65:4,190:5,100:5,226:3,243:1,154:5,214:4,161:4,62:3,188:3,102:2,69:3,170:5,38:3,9:5,246:5,22:4,21:1,179:3,187:4,135:4,68:4,146:4,176:5,166:5,138:1,247:1,89:5,2:3,30:3,63:2,249:4,269:5,32:5,141:3,211:3,40:3,270:5,133:4,239:4,194:4,256:4,220:3,93:5,8:1,205:3,234:4,105:2,147:3,99:3,1:5,197:5,173:5,75:4,268:5,34:2,144:4,271:2,119:5,26:3,158:3,37:2,181:5,136:3,257:4,237:2,131:1,109:5,182:4,71:3,223:5,46:4,169:5,41:2,162:4,110:1,66:4,77:4,199:4,57:5,50:5,192:4,178:5,5:3,87:5,238:4,156:4,106:4,167:2,115:5,11:2,245:2,35:1,137:5,127:5,16:5,79:4,261:1,45:5,48:5,25:4,251:4,195:5,153:3,101:2,168:5,123:4,191:5,4:3,263:1,203:4,55:5,42:5,139:3,240:3,7:4,149:2,43:4,165:5,116:3,198:5,124:5,95:4,217:3,58:4,142:2,216:5,126:2,83:3,231:1,204:5,3:4,207:5,244:2,19:5,29:1,18:4,59:5,15:5,111:5,52:4,88:4,13:5,28:4,172:5,122:3,152:5,94:2

Now the recommendation system put forth possible recommendations for user 'u1' by observing the recommendation given for other items by 'u1'. The various recommendations against the user specified threshold values are shown in the Table 2 and the graphical representation has been presented in Figure 4.

Table 2: Number of items recommended for the Active User '1' By varying recommended threshold values

Recommended threshold	*Recommended Movie id with rating*
10	[11:5.0,22:5.0,7:5.0,10:5.0,18:5.0,20:5.0,2:5.0,4:5.0,8:5.0,23:5.0]
20	[25:5.0,38:5.0,18:5.0,24:5.0,30:5.0,37:5.0,7:5.0,11:5.0,22:5.0,23:5.0,26:5.0,28:5.0,31:5.0,33:5.0,2:5.0,4:5.0,8:5.0,10:5.0,20:5.0,39:5.0]
50	[58:5.0,88:5.0,30:5.0,56:5.0,72:5.0,83:5.0,18:5.0,28:5.0,43:5.0,54:5.0,68:5.0,71:5.0,76:5.0,80:5.0,7:5.0,11:5.0,23:5.0,26:5.0,37:5.0,40:5.0,49:5.0,53:5.0,65:5.0,67:5.0,69:5.0,70:5.0,73:5.0,75:5.0,77:5.0,79:5.0,2:5.0,4:5.0,8:5.0,10:5.0,20:5.0,22:5.0,24:5.0,25:5.0,31:5.0,33:5.0,38:5.0,39:5.0,46:5.0,47:5.0,51:5.0,52:5.0,61:5.0,62:5.0,66:5.0,90:5.0]

Recommended threshold	*Recommended Movie id with rating*
100	[122:5.0,180:5.0,58:5.0,121:5.0,148:5.0,179:5.0,30:5.0,56:5.0,79:5.0,118:5.0,139:5.0,147:5.0,158:5.0,167:5.0,18:5.0,28:5.0,43:5.0,54:5.0,69:5.0,77:5.0,95:5.0,117:5.0,133:5.0,138:5.0,143:5.0,146:5.0,153:5.0,157:5.0,161:5.0,164:5.0,7:5.0,11:5.0,23:5.0,26:5.0,37:5.0,40:5.0,49:5.0,53:5.0,65:5.0,68:5.0,72:5.0,76:5.0,88:5.0,94:5.0,99:5.0,116:5.0,126:5.0,132:5.0,135:5.0,136:5.0,141:5.0,142:5.0,144:5.0,145:5.0,149:5.0,151:5.0,155:5.0,156:5.0,159:5.0,160:5.0,162:5.0,163:5.0,2:5.0,4:5.0,8:5.0,10:5.0,20:5.0,22:5.0,24:5.0,25:5.0,31:5.0,33:5.0,38:5.0,39:5.0,46:5.0,47:5.0,51:5.0,52:5.0,61:5.0,62:5.0,66:5.0,67:5.0,70:5.0,71:5.0,73:5.0,75:5.0,80:5.0,83:5.0,90:5.0,92:5.0,97:5.0,98:5.0,102:5.0,107:5.0,123:5.0,125:5.0,130:5.0,131:5.0,134:5.0,182:5.0]

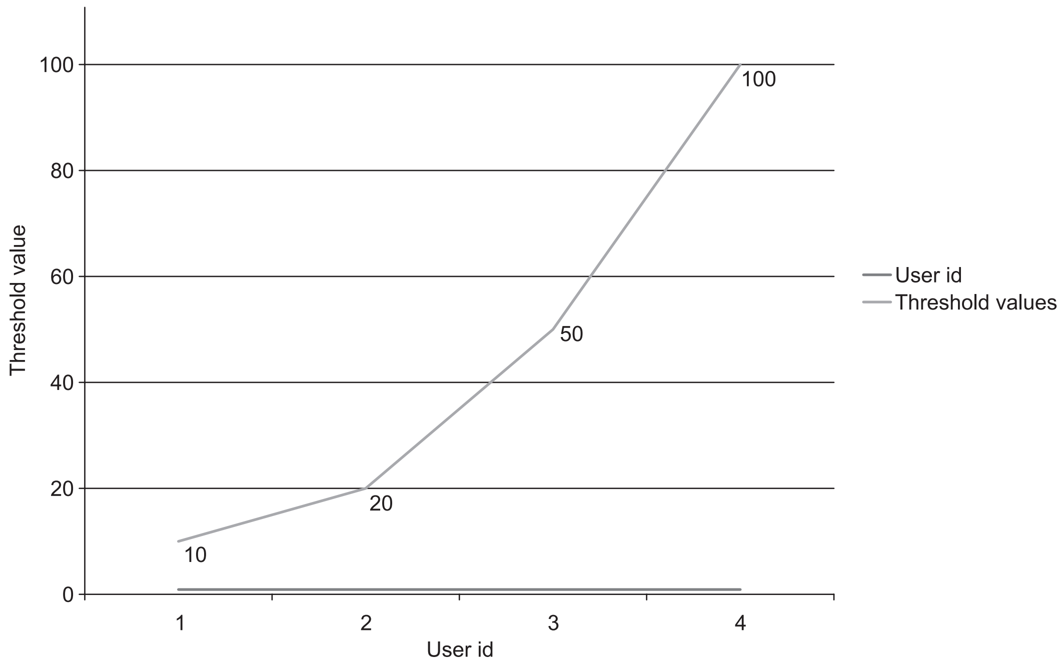

Figure 4: Number of items recommended for the Active User '1' By varying recommended threshold values

Similarly the movie recommendations for all the users those who are not specified their recommendations for various items have been calculated. The proposed recommended movie rating along with the movie-id for the 25 users have been calculated and presented in Table 3.

Table 3: Top 'N' Recommendation for various user where 'N' = 10

User id	*Movie id & Rating*
1	[11:5.0,22:5.0,7:5.0,10:5.0,18:5.0,20:5.0,2:5.0,4:5.0,8:5.0,23:5.0]
2	[7:5.0,11:5.0,2:5.0,6:5.0,8:5.0,9:5.0,1039:5.0,1:5.0,4:5.0,12:5.0]
3	[1226:5.0,130:5.0,128:5.0,179:5.0,1144:5.0,1138:5.0,1127:5.0,1143:5.0,1170:5.0,1231:5.0]
4	[12:5.0,25:5.0,5:5.0,11:5.0,13:5.0,22:5.0,1:5.0,2:5.0,7:5.0,28:5.0]
5	[7:5.0,11:5.0,4:5.0,6:5.0,8:5.0,9:5.0,1:5.0,2:5.0,5:5.0,12:5.0]
6	[8:5.0,11:5.0,4:5.0,7:5.0,9:5.0,10:5.0,1:5.0,2:5.0,6:5.0,12:5.0]
7	[10:5.0,14:5.0,3:5.0,6:5.0,11:5.0,13:5.0,1:5.0,2:5.0,5:5.0,15:5.0]
8	[7:5.0,12:5.0,2:5.0,1553:5.0,8:5.0,11:5.0,1039:5.0,1:5.0,4:5.0,13:5.0]
9	[14:5.0,17:5.0,2:5.0,569:5.0,15:5.0,1478:5.0,150:5.0,1115:5.0,564:5.0,20:5.0]
10	[8:5.0,13:5.0,4:5.0,7:5.0,9:5.0,11:5.0,1:5.0,2:5.0,5:5.0,16:5.0]

User id	*Movie id & Rating*
11	[11:5.0,14:5.0,4:5.0,7:5.0,12:5.0,13:5.0,1:5.0,2:5.0,6:5.0,18:5.0]
12	[8:5.0,12:5.0,3:5.0,7:5.0,9:5.0,11:5.0,1:5.0,2:5.0,5:5.0,13:5.0]
13	[7:5.0,10:5.0,3:5.0,6:5.0,8:5.0,9:5.0,1:5.0,2:5.0,5:5.0,11:5.0]
14	[8:5.0,13:5.0,3:5.0,6:5.0,9:5.0,11:5.0,1:5.0,2:5.0,4:5.0,14:5.0]
15	[6:5.0,12:5.0,2:5.0,4:5.0,1046:5.0,1047:5.0,303:5.0,1:5.0,1042:5.0,1053:5.0]
16	[17:5.0,24:5.0,5:5.0,16:5.0,20:5.0,23:5.0,2:5.0,3:5.0,13:5.0,25:5.0]
17	[762:4.6,1327:4.6,1170:4.5,1134:4.5,1281:4.5,525:4.5,281:4.5,117:4.5,593:4.5,1060:4.5]
18	[11:5.0,17:5.0,5:5.0,10:5.0,15:5.0,16:5.0,2:5.0,4:5.0,7:5.0,18:5.0]
19	[162:5.0,234:5.0,657:5.0,132:5.0,141:5.0,58:4.8,1479:4.7 ,512:4.7,614:4.7,500:4.6]
20	[722:5.0,686:5.0,662:5.0,657:5.0,648:5.0,534:5.0,146:5.0,1058:5.0,1501:5.0,939:4.7]
21	[12:5.0,17:5.0,5:5.0,11:5.0,13:5.0,15:5.0,2:5.0,4:5.0,8:5.0,19:5.0]
22	[8:5.0,13:5.0,3:5.0,7:5.0,11:5.0,12:5.0,1:5.0,2:5.0,5:5.0,16:5.0]
23	[11:5.0,14:5.0,6:5.0,9:5.0,12:5.0,13:5.0,2:5.0,4:5.0,7:5.0,16:5.0]
24	[6:5.0,14:5.0,3:5.0,5:5.0,7:5.0,13:5.0,1:5.0,2:5.0,4:5.0,15:5.0]
25	[12:5.0,18:5.0,1553:5.0,11:5.0,13:5.0,15:5.0,2:5.0,4:5.0,7:5.0,21:5.0]

4. CONCLUSION

The traditional information retrieval methods are alone not enough for analysing the casual relationship among retrieved objects. Besides, the casual and intensive relationships among the objects are to be utilized for the task of personalization. With the advent of big data, information has been flown from diversified data sources which are highly independent in nature. To reap meaningful pattern of interest as part of big data mining task, a distributed machine learning framework for item-based collaborative filtering has been implemented in this paper. Besides the successful online recommendation systems and their elevation towards big data environment is also highlighted. Implementation of effective similarity measures, provision of hybrid recommendation system on the basis improved techniques would be some of the possible future work from our end.

REFERENCES

[1] Ramkumar T, Hariharan S, Selvamuthukumaran S., A survey on mining multiple data sources, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2013, 3(1):1-11.

[2] Ramkumar T, Srinivasan R., Multi-Level Synthesis of Frequent Rules from Different Data-Sources,International Journal of Computer Theory and Engineering, 2010, 2(2):195-204.

[3] Ramkumar T, Srinivasan R., The Effect of Correction Factor in Synthesizing Global Rules in a Multi-databases Mining Scenario, Journal of Applied Computer Science & Mathematics, 2009,3(6) :33-38.

[4] Ramkumar T, Srinivasan R, Hariharan S., Synthesizing global association rules from different data sources based on desired interestingness metrics, International Journal of Information Technology and Decision Making, 2014, 13(3):473-495.

[5] Hariharan S, Ramkumar T., Mining product reviews in web forums, Information Retrieval Methods for Multidisciplinary Applications, 2013, 78-94.

[6] Bazeer Ahamed B, Ramkumar T., An intelligent web search framework for performing efficient retrieval of data, Computers and Electrical Engineering, 2016, 56:289-299.

[7] Goldberg D, Nichols D, Oki BM, Terry D., Using collaborative filtering to weave an information tapestry. Communications of the ACM, 1992, 35(12):61-70.

[8] Adomavicius G, Tuzhilin A., Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions, IEEE transactions on knowledge and data engineering, 2005,17(6):734-49.

[9] Chen LS, Hsu FH, Chen MC, Hsu YC. Developing recommender systems with the consideration of product profitability for sellers, Information Sciences, 2008, 178(4):1032-48.

[10] Ziegler C.N., McNee S.M., Konstan J.A. and Lausen G., Improving recommendation lists through topic diversification. In Proceedings of the 14th international conference on World Wide Web, 2005, 22-32.

[11] Mooney, Raymond J., and Loriene Roy, Content-based book recommending using learning for text categorization, In Proceedings of the 5th ACM conference on Digital libraries, 2000, 195-204.

[12] Billsus D, Pazzani MJ., User modeling for adaptive news access. User modeling and user-adapted interaction, 2000, 10(2-3): 147-80.

[13] Vesin B, Ivanović M, KlašNja-MilićEvić A, Budimac Z., Protus 2.0: Ontology-based semantic recommendation in programming tutoring system. Expert Systems with Applications. 2012, 39(15):12229-46.

[14] Konstas I, Stathopoulos V, Jose JM.,On social networks and collaborative recommendation. In Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, 2009, 195-202.

[15] Cunningham P, Bergmann R, Schmitt S, Traphöner R, Breen S, Smyth B.,Websell: Intelligent sales assistants for the world wide web, 2001,15(1):28-32.

[16] Ghazanfar MA, Prugel-Bennett A., A scalable, accurate hybrid recommender system, Third International Conference Knowledge Discovery and Data Mining, pp. 94-98.

[17] Sarwar BM, Konstan JA, Borchers A, Herlocker J, Miller B, Riedl J., Using filtering agents to improve prediction quality in the grouplens research collaborative filtering system. In Proceedings of the 1998 ACM conference on Computer supported cooperative work 1998, 345-354.

[18] Zhao ZD, Shang MS., User-based collaborative-filtering recommendation algorithms on hadoop. 3rd International conference on Knowledge Discovery and Data Mining, 2010, 478-481.

[19] De Pessemier T, Vanhecke K, Dooms S, Martens L., Content-based recommendation algorithms on the hadoop mapreduce framework, 7th International Conference on Web Information Systems and Technologies, 2011, 237-240.