Proceedings of 4th RIT Post Graduates Conference (RIT PG Con-18)
NOVATEUR PUBLICATIONS
JournalNX- A Multidisciplinary Peer Reviewed Journal (ISSN No:2581-4230)
April, 13th, 2018

SOCIAL GROUP RECOMMENDATION SYSTEM BASED ON BIG DATA

Ms. Nikita S. Mohite
Department of Technology Shivaji University Kolhapur, India
email:mohitenikita1995@gmail.com

Mr. H. P. Khandagle
Department of Technology. Shivaji University Kolhapur, India
email:k_hriday@yahoo.com

***Abstract*—In recent years, the use of internet and functional activities have created development to evolve the system and application of Cyber-Physical-Social Systems (CPSSs).Cyber-Physical-Social Systems (CPSSs) became the essential criteria of evolution within the data business, through that ancient technology can evolve into cyber-physical-social process science. Existing work is recommended person for example Facebook. This project, proposes a web based application on multidimensional system that the group-centric recommender system within the CPSCP domain with activity-oriented cluster discovery, the revision of rating information for improved accuracy, and cluster preference modelling that supports decent context mining from multiple sources. To boot we have a tendency to inserting additional four-dimensional cluster preference, modelling like profile primarily based, content primarily based. In profile based profile is going to be refer and content in content based. The goal of all over system is to study and development with specific techniques and methods for obtaining user references from several interactions with the group member objective to make the system. The - recommender system is economical, objective and correct.**

***Keywords*— Big data, Data mining, Context mining, Big data analysis, Cluster preferences modelling, Recommended services, CPSCPs.**

I. INTRODUCTION

Basically these days Internet of things are popular and widely used and more attractive. Because of Physical perceptions, Cyber interactions, Social correlations and even cognitive things which are interconnected. On Internet online group activities are increased to communicate with the need for a group recommendation system to become more and more mandatory. In past years computing, communication, control, management, information development, etc. has been successfully applied to many important systems like healthcare, industries, social media, manufacturing, defense and dramatically increases their controllability, adaptability, autonomy, efficiency, functionality, reliability, safety, usability and main is the functionality.

Three important things in designing CPSS based transportation are: The physical system that is proposed by physics of transportation system. The cyber system that is proposed by distributed computing and communications and social system proposed by human organization and behavioural decision making. From the research paper, it comes to know that CPSS must be conducted with a multidisciplinary approach including the physical, social and cognitive science and that AI based system is key to any successful construction and deployment [12], [13], [15].

The Social Network Services (SNSs) is used to increasing amounts of information from CPSS being generated and disseminated through social networks [6].

To storing the data which is fulfil by new user/user it store by using following techniques:-

1. Apache Hadoop:-

Apache Hadoop is a java based free software framework that can effectively store large amount of data in a cluster [11]. This framework runs in parallel on a cluster and has an ability to allow us to process data across all nodes. Hadoop Distributed File System (HDFS) is the storage system of Hadoop which splits big data and distribute across many nodes in a cluster. This also replicates data in a cluster thus providing high availability.

2. NoSQL:-

While the traditional SQL can be effectively used to handle large amount of structured data, we need NoSQL (Not Only SQL) to handle unstructured data. NoSQL databases store unstructured data with no particular schema. Each row can have its own set of column values. NoSQL gives better performance in storing massive amount of data. There are many open-source NoSQL DBs available to analyse big Data.

3. Hive:-

This is a distributed data management for Hadoop. This supports SQL-like query option HiveSQL (HSQL) to access big data. This can be primarily used for Data mining purpose. This runs on top of Hadoop.

4. Sqoop:-

This is a tool that connects Hadoop with various relational databases to transfer data. This can be effectively used to transfer structured data to Hadoop or Hive.

5. Presto:-

Facebook has developed and recently open-sourced its Query engine (SQL-on-Hadoop) named presto which is built to handle petabytes of data. Unlike Hive, Presto does not depend on Map Reduce technique and can quickly retrieve data.

II. NEED OF WORK

This preference-oriented social networks with the power to get cluster choices or recommendations even once the preferences of some cluster members area unit unobserved. Most cluster recommendation approaches serve teams that encompass permanent members; but, their accuracy is also suffering from variations in

individual preference. Particularly, it's universally accepted that cluster recommendation is a lot of economical and effective than individual recommendation. The existing approaches are still supported with low-dimensional rating knowledge that are not directly obtainable to be used in CPSSs. All existing emotion-aware recommender systems concentrate on the emotional necessities of the user.

The purpose of this system is to communication with social media and share users post. So the SNS platform is used in the CPSS. Physical & social system and their cyber systems can understand each other, their stepwise interaction helps improve each other, and gradually the effective control and safe, reliable and efficient operation of CPSS will be realized.

SNS and CPSS integrated into our proposed system named as CPSCPs (Cyber-Physical-Social-Content-Profile based system) in which content based system content is going to be refer and profile is refer in profile based system.

III. SYSTEM ARCHITECTURE

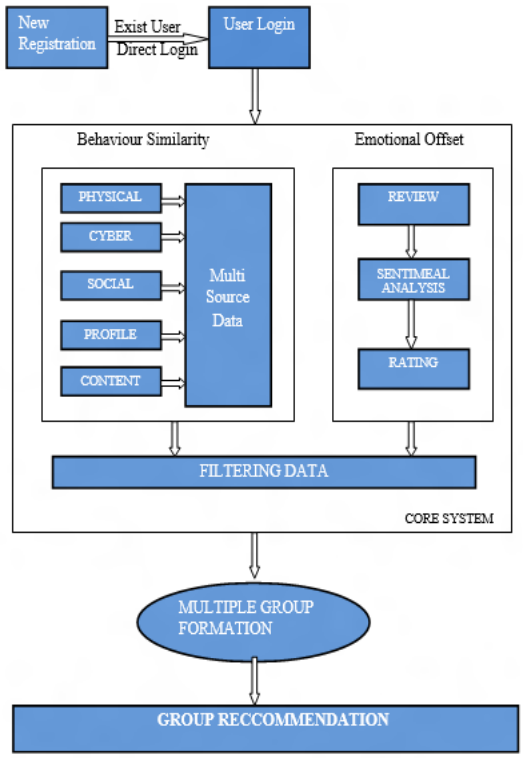

Fig.1: Architecture Diagram of Proposed Work

This system is going to be developed as a creative website in which Group Recommendation is a main approach. This Group Recommendation is based on behaviour similarity of the user and emotional offset. To making the group on the basis of behaviour and content based using multidimensional database storage i.e. rollup and cube for similar properties and similar attributes of people.

Behaviour similarity invokes the attributes like cyber, physical, social, profile and content. These make multidimensional data which improve the challenge of individual centric recommendation. In emotional offset, it's based on reviews and rating. At a time of sentimental analysis the topic searching part from user reviews are done. LDA is used in sentimental analysis for finding topic from users post. Emotional features are extracted from sentimental analysis the rating module improve the accuracy of recommendation result.

IV. MODULES

The proposed system will be consist of 6 modules:-

1. Login Module:-

Login module is a web application where user can login. If user is existing user then they can login with the help of user-id and password. If user don't have the authentication then user click on new user registration.

2. New User Registration:-

It is necessary to register the new user and get the user name and password for login. In this registration form person fill the all personal information like name, DOB, age, current location, school name, etc.

3. Core Module:-

Core module consist of Big data for efficient storage of multidimensional data.

a) Group Discovery supported behavioural Similarity:-

In Group Recommendation, the cluster discovery module is meant to speedily and accurately discover teams from high-dimensional information supported the similarity of user behaviour**.**

b) Method of Rating Revision supported Emotional Offset selection through LDA:-

In Group Recommendation, rating information are revised in accordance with their emotional options that are extracted from user reviews through sentiment analysis. Considering that user reviews will embrace a major emotional offset, sentiment analysis is a crucial suggests that of calculative this offset to revise the initial ratings. The rating revision module is that the foundation for guaranteeing the objectiveness of the recommender, thereby raising the accuracy of the advice results. LDA (Latent Diritchlet Allocation) is used in sentimental analysis for finding the topic from posts. While doing the sentimental analysis the topic searching part from user reviews are done. This will help for the formation of group based on reviews.

4. Filtering Of Data:-

In the filtering of data format we filter all cyber-physical-social-content-profile and make in standard format with similarities and store again on our database for making the group with similar data. For this purpose we will use KNN, K-means, collaborative filtering.

5. Multiple Group Formation:-

To making the group on the basis of behavioural and content base using rollup and cube for similar properties and similar attributes of people.

6. Group Recommendation:-

Whenever new user login then their activities are matched with our existing groups. Then our system will be sending recommendation for that person to add it.

V. METHODOLOGY

1 Methods of data collection

We can collect the data through web application.

Web Application:-

We are designing Web application for data collection with personal details like college name, location, working organization, daily activities etc. We are using Facebook, Twitter for collecting data and there is also new registration form from which we will get standard format.

2 Probable methods of data analysis

We use the NoSQL (Not Only SQL) to handle unstructured data. NoSQL databases store unstructured data with no particular schema.

Social information area unit is a vital supplement to context analysis during a recommender system, they embrace varied styles of emotional data that influence the judgment of the advice results. With the speedy development of SNSs, varied approaches integrated with social information are projected. The projected approach will solely be applied to research information during specific contexts, that don't seem to be on the market in CPSSs. We tend to propose to represent user behaviour information in tensor kind, analyse the activity similarity of users via Tucker decomposition, and see teams via agglomeration.

The collaborative filtering system used as background data rating performed by the user. As input data is stored the rating from the actual user has done about one or more items of the system. The recommenders look for the similarities and make suggestion to make group of similar rating. In collaborative filtering item can consist of anything for which a human can provide rating such as art, books, CDs, journals, articles or vacation destination.

This method is based on collectively and analysing large amount of information on user behaviour activities preferences and predicting what user will like based on their similarities to other users and recommend the group to those recommend users [10].

3. Method for multidimensional data storage

Multidimensional database enables interactive analyses of large amounts of data for decision-making purposes.

The k Nearest Neighbors Algorithm (KNN)

The Algorithm

1. A positive integer k is specified, along with a new sample.
2. We select the k entries in our database which are closest to the new sample.
3. We find the most common classification of these entries.
4. This is the classification we give to the new sample.

****END KNN****

Simple KMeans Clustering

KMeans is an iterative clustering algorithm in which items are moved among set of clusters until the desired set is reached. This can be viewed as a type of squared error algorithm. In almost all cases, the simple KMeans clustering algorithm [18] takes more time to form clusters. So it is not suitable to be employed for large datasets.

Farthest First Clustering

Farthest first is a variant of K Means. This places the cluster center at the point further from the present cluster. This point must lie within the data area. The points that are farther are clustered together first. This feature of farthest first clustering algorithm speeds up the clustering process in many situations like less reassignment and adjustment is needed. But in this system more reassignment and no adjustment is there so this method is not applicable here.

Make Density Based Clustering

A cluster is a dense region of points that is separated by low density regions from the tightly dense regions. This clustering algorithm can be used only when the clusters are irregular.

K-means clustering algorithm

The *k*-means method has been shown to be effective in producing good clustering results for many practical applications**.** K-means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. A simple and efficient implementation of efficient K-Means clustering algorithm called the filtering algorithm shows that the algorithm runs faster as the separation between clusters increases [19].

VI. RESULT ANALYSIS

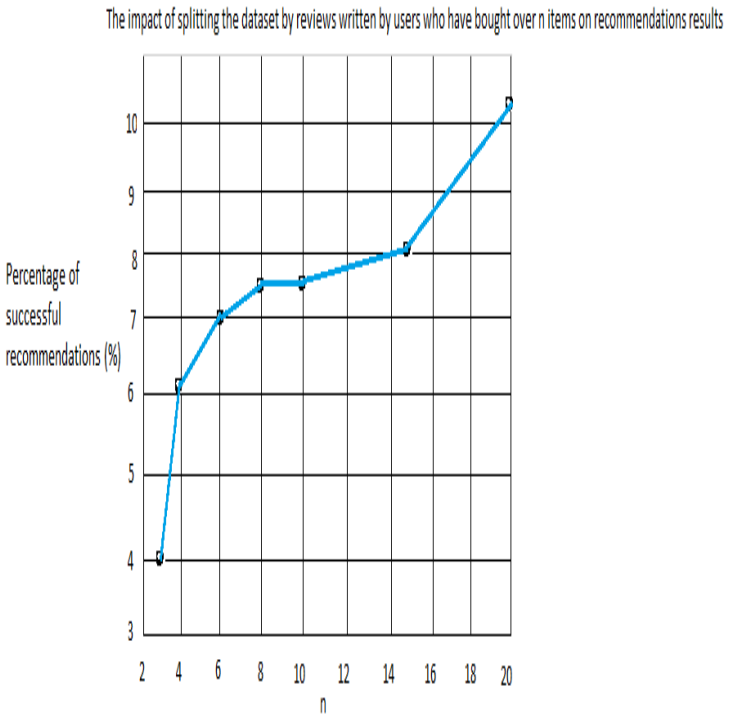

Fig.2 The Impact of splitting the dataset by reviews written by users who have join group over on user the recommendation results**.**

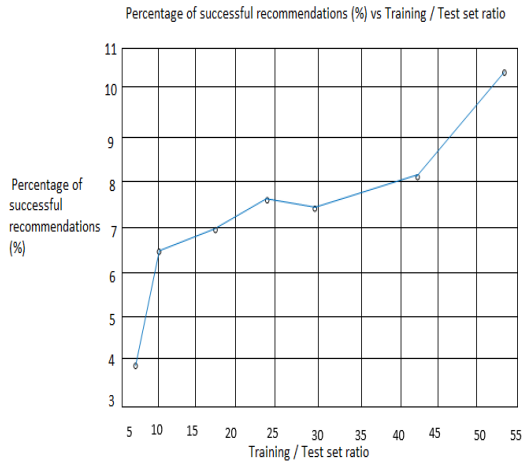

Fig.3 Percentage of successful recommendations (%) vs. Training/Test set ratio.

Percentage of Successful Recommendations (%) vs. Training/Test Set Ratio

Figure 2 shows the impact of splitting the dataset by reviews written by users who have join over n group on the recommendation results. Figure 3 shows the impact of training/test set ratio on the recommendation results. Figure 2 and figure 3 are the same because each one particular n corresponds to one particular training/test set ratio. From the two figures, we can see that as n and training/test set ratio increase, the percentage of successful recommendations increases because the more user the users have joined group, the more correlated information we can obtain to give recommendations.

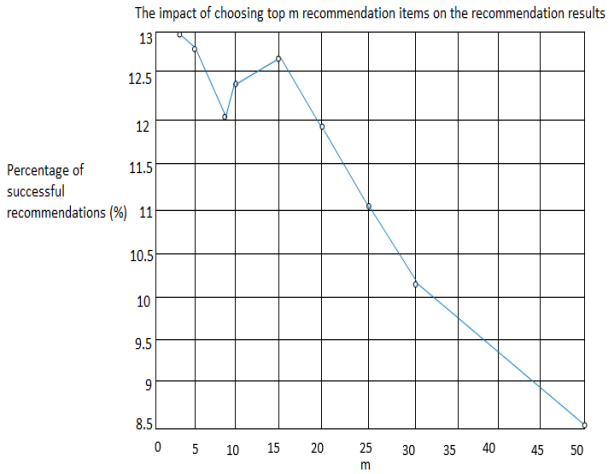

Fig.4 The impact of choosing top m recommendation group on the recommendation results.

Percentage of Successful Recommendations (%) vs. Top m Recommendation items

Fig.4 show the impact of choosing top m recommendation group on the recommendation results. The figure, it can be found that when we choose m smaller than 15, the percentage of successful recommendations remains higher than 12%, which indicates good recommendations. As m increases and becomes larger than 20, the percentage of successful recommendations decreases. In the reality, we should always recommend suitable number of group to the potential users. If we recommend too many groups, then uer will feel bored and show no interest in even glancing at them. On the other hand, if we recommend too few group, the users will not have enough choices. As a result, a suitable number of recommendation group should be selected carefully for recommendation system in reality.

VII. CONCLUSION

This system is going to be developed as a creative website in which Group Recommendation is a main approach. This Group Recommendation is based on behaviour similarity of the user and emotional offset. SNS provides individual centric recommendation, this make complex, non-efficient system. So to come over this problem we comes with Group centric recommendation based on CPSCPs. This provide effective, objective and accurate recommendation services in CPSCPs. This help to reduce complexity of conventional individual centric recommender systems. The emotional offset extracted from reviews, sentiment analysis is proposed to revise user rating for improved result of rating data. Cluster data is achieve by applying KNN, K-means, collaborative filtering, content based filtering. This group recommendation system gives efficient, effective and accurate results. At the end it will give recommendation of the group that will help in various field like professional field, Industrial field, and personal day-to-day life.

Acknowledgment

This review paper work was guided and supported by Mr. H. P. Khandagle. I would like to thank the guide anonymous reviewers for their valuable and constructive comments on improving the paper.

References

[1] Alexander W. Schneider and Boltzmannstr, “Using Web Analytics Data to support Social Software Users”, 2010, srvmatthes5.in.tum.de.
[2] Gantz, J., & Reinsel, D. (2011). The 2011 Digital Universe Study: Extracting Value from Chaos.
[3] Internet source, aisel.aisnet.org
[4] A. Sheth, P. Anantharam, and C. Henson, “Physical-cyber-social computing: An early 21st century approach,” Intelligent Systems, IEEE, vol. 28, no. 1, pp. 78–82, 2013.
[5] Y.-L. Chen, L.-C. Cheng, and C.-N. Chuang, “A group recommendation system with consideration of interactions among group members,” Expert systems with applications, vol. 34, no. 3, pp. 2082– 2090, 2008.
[6] Gautam Shroff, Lipika Dey and Puneet Agrawal, “Social Business Intelligence Using Big Data”, 2013.
[7] V. Jude Nirmal and D.I. George Amalarethinam, “Parallel Implementation of Big Data Pre-Processing Algorithms for Sentiment Analysis of Social Networking Data”, IJFMA Vol. 6, No. 2,

2015, 149-159, ISSN: 2320 –3242 (P), 2320 – 3250 (online) Published on22 January 2015.

[8] R. Balaji Ganesh & S. Appavu, "An Intelligent Video Surveillance Framework with Big Data Management for Indian Road Traffic System", IJCA (0975 – 8887) Volume 123 – No.10, August 2015.

[9] Feng Chen, Pan Deng, Jiafu Wan, Daqiang Zhang, Athanasios V. Vasilakos and Xiaohui Rong, "Data Mining for the Internet of Things: Literature Review and Challenges", Hindawi Publishing Corporation, International Journal of Distributed Sensor Networks, Article ID 431047, March 2015.

[10] Somesh S Chavadi and Dr. Asha T, "Text Mining Approach for Big Data Analysis Using Clustering and Classification Methodologies", IJETAE, Volume 4, Issue 8, August 2014.

[11] Vatsal Shah, "Big Video Data Analytics using Hadoop", IJARCSSE, Vol 5, issue 7, July 2015.

[12] F.-Y. Wang, "The emergence of intelligent enterprises: From CPS to CPSS," Intelligent Systems, IEEE, vol. 25, no. 4, pp. 85–88, 2010.

[13] G. Xiong, F. Zhu, X. Liu, X. Dong, W. Huang, S. Chen, and K. Zhao, "Cyber-physical-social system in intelligent transportation," Auto-matica Sinica, IEEE/CAA Journal of, vol. 2, no. 3, pp. 320–333, 2015.

[14] Sana Siddiqui and Imran Qadri, "Mining Web Log Files for Web Analytics and Usage Patterns to Improve Web Organization", Siddiqui et al., IJARCSSE 4(6), June - 2014, pp. 794-802.

[15] Y. Hu, F.-Y. Wang, and X. Liu, "A CPSS approach for emergency evacuation in building fires," IEEE Intelligent Systems, no. 3, pp. 48–52, 2014.

[16] M. Chen, S. Mao, and Y. Liu, "Big data: A survey,"Mobile Networks and Applications, vol. 19, no. 2, pp. 171–209, 2014.

[17] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," in Proceedings of the fourth ACM conference on Recommender systems. ACM, 2010, pp. 135–142.

[18] Shi Na, L. Xumin, G. Yong, "Research on K-Means clustering algorithm-An Improved K-Means Clustering Algorithm", "IEEE Third International Symposium on Intelligent Information Technology and Security Informatics", pp.63-67, Apr.2010.

[19] Tapas Kanungo, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu, "An Efficient k-Means Clustering Algorithm: Analysis and Implementation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, VOL. 24, NO. 7, PP. 881-892,2000.