Advanced Recommendation System

Rohit Bhanushali[1], Smeet Chitalia[2], Prem Panchal[3]

[123]Student, Dept. of Computer Engineering, K.J Somaiya Institute of Engineering and Information Technology, Sion, Mumbai-22, Maharashtra, India

---***---

Abstract - *Recommendation System has become more and more popular in recent years. To facilitate the purchase process, many online stores provide a shopping recommendation system for their consumers. So far, the generic recommendation systems mainly consider preferences of a consumer based on his/her purchase history. But recommendation based only on consumer's purchase history is a major drawback. If a consumer wishes to buy a product that he/she never bought then that recommendation system will fail. Our system tries to overcome this drawback. It not only checks purchase history but various other parameters come into play. Our system recommends products, music and movies based on location, browsing history of the users, ratings from different users and highly purchased products. System will recommend similar types of products for people having similar buying trends. It aims at supporting the users in various decision making processes such as what items to buy, what music to listen, which movie to watch, etc. In our new algorithm, we retrieve live ratings, data of news and music from Twitter. We believe that such a new scheme should be able to provide a better recommendation list which fit consumer's desire.*

Key Words: **Recommendation, Collaborative Filtering, Clustering, Twitter, Classifier, Analysis, Product, Music, News.**

1. INTRODUCTION

Most online stores provide a shopping recommendation system for the consumers to facilitate online shopping. The core of such systems is a personalized recommendation algorithm. This algorithm models consumer shopping behaviors and recommend items to the consumers while doing on-line purchasing. Since there is no explicit product rating available for shopping, the system has to estimate consumers' preferences from their purchased histories. One of the major techniques used to develop a recommendation algorithm is collaborative filtering (CF). Nevertheless, the problem of inadequacy due to too few user ratings may make the formation of neighborhood inaccurate and there by results in poor recommendations.

1.1 Recommendation

- The act of saying that someone or something is good and deserves to be chosen.
- A suggestion about what should be done.
- A suggestion or proposal as to the best course of action, especially one put forward by an authoritative body.

1.2 Collaborative Filtering

Collaborative Filtering (CF) is a technique used by some recommender systems. Collaborative filtering has two senses, a narrow one and a more general one. In the newer, narrower sense, collaborative filtering is a method of making automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many users (collaborating).

- **Workflow of Collaborative Filtering:**

1. A user expresses his/her preferences by rating items of the system. These ratings can be viewed as an appropriate representation of the user's interest in the corresponding domain.
2. The system matches this user's ratings against other users' and finds the people with most "similar" tastes.
3. With similar users, the system recommends items that the similar users have rated highly but not yet being rated by this user (presumably the absence of rating is often considered as the unfamiliarity of an item).

1.3 Clustering

Clustering or Cluster Analysis is the task of grouping a set of objects in such a way that objects in the same group (called a **cluster**) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning.

- Clustering is a process of partitioning a set of data (or objects) into a set of meaningful sub-classes, called clusters.
- Help users understand the natural grouping or structure in a data set.
- Clustering: unsupervised classification: no predefined classes.
- Used either as a stand-alone tool to get insight into data distribution or as a pre-processing step for other algorithms.
- Moreover, data compression, outlier's detection, understands human concept formation.

2. LITERATURE REVIEW

Li et al. proposed to incorporate multidimensional clustering into a collaborative filtering recommendation model. Background data in the form of user and item profiles was collected and clustered using the proposed algorithm in the first stage. Then the poor clusters with similar features were deleted while the appropriate clusters were further selected based on cluster pruning. At the third stage, an item prediction was made by performing a weighted average of deviations from the neighbor's mean. Such an approach was likely to trade-off on increasing the diversity of recommendations while maintaining the accuracy of recommendations. [1], [4]

Zhou et al. represented Data-Providing (DP) service in terms of vectors by considering the composite relation between input, output, and semantic relations between them. The vectors were clustered using a refined fuzzy C-means algorithm. Through merging similar services into a same cluster, the capability of service search engine was improved significantly, especially in large Internet based service repositories. However, in this approach, it is assumed that domain ontology exists for facilitating semantic interoperability. Besides, this approach is not suitable for some services which are lack of parameters. [2]

Simon et al. used a high-dimensional parameter free, divisive hierarchical clustering algorithm that requires only implicit feedback on past user purchases to discover the relationships within the users. Based on the clustering results, products of high interest were recommended to the users. However, implicit feedback does not always provide sure information about the user's preference. [3]

3. EXISTING SYSTEM

In the existing system with the prevalence of service computing and cloud computing, more and more services are deployed in cloud infrastructures to provide rich functionalities. Service users have nowadays encounter unprecedented difficulties in finding ideal ones from the overwhelming services. Recommender systems (RSs) are techniques and intelligent applications to assist users in a decision making process where they want to choose some items among a potentially overwhelming set of alternative products or services. Collaborative filtering (CF) such as item- and user-based methods are the dominant techniques applied in RSs.

The most fundamental challenge for the Big Data applications is to explore the large volumes of data and extract useful information or knowledge for future actions. The basic assumption of user-based CF is that people who agree in the past tend to agree again in the future.

Different with user-based CF, the item-based CF algorithm recommends a user the items that are similar to what he/she has preferred before in traditional CF algorithms, to compute similarity between every pair of users or services may take too much time, even exceed the processing capability of current RSs. Consequently, service recommendation based on the similar users or similar services would either lose its timeliness or could not be done at all.

- **Disadvantages of existing system:**

1. Existing System making decision is time consuming.
2. The cluster analysis gathers users with similar characteristics according to the web visiting message data only.
3. Much Time to Search and data Clustering Management is poor performance.
4. Through merging similar services into a same cluster, especially in large Internet-based service repositories. This approach is not suitable for some services.
5. In this system use a specific one type of recommendation technique so the recommendation is poor.

4. PROPOSED SYSTEM

We propose a Clustering-based Collaborative Filtering approach, which consists of two stages: clustering and collaborative filtering. Clustering is a pre-processing step to separate big data into manageable parts. A cluster contains some similar services just like a club contains some like-minded users. Since the number of services in a cluster is much less than the total number of services, the computation time of CF algorithm can be reduced significantly. Besides, since the ratings of similar services within a cluster are more relevant than that of dissimilar services, the recommendation accuracy based on users' ratings may be enhanced. In this system fetch social media data, analysis by using the collaborative filtering algorithm for getting specific output and using this data give personal recommendation for consumer.

Following modules are developed:

1. User History: The browsing history of user, so that user interest can be found.
2. Based on popularity: Popular product that are that easily available and users ask accesses/views it frequently.
3. Collaborative filtering: Collaborative algorithm will be applied on the basis of product rating on the above data and then product will be shown to user.
4. Social media (Twitter):-Trend on social media will be captured and will be recommended.
5. Personalized recommendation:-Based on user personal information product will be displayed to user.

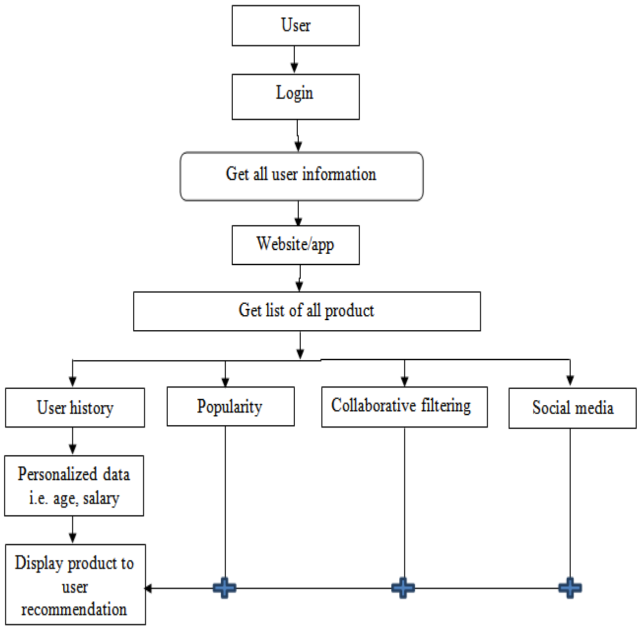

Fig -1: System Architecture

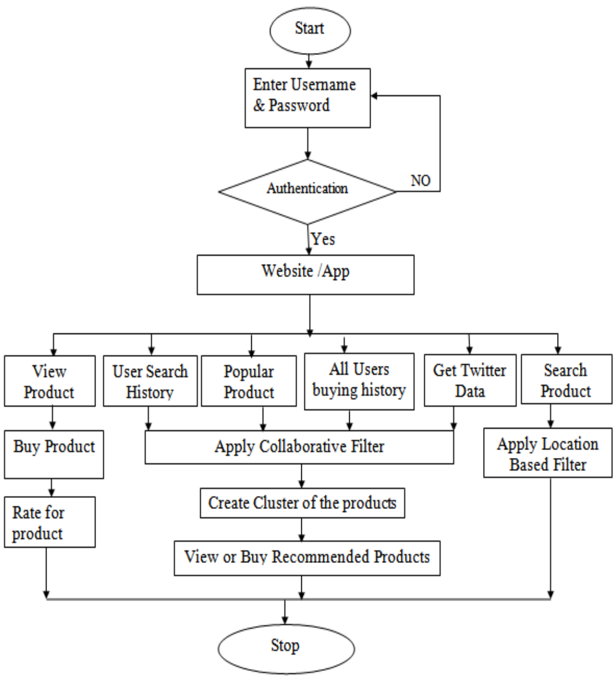

Fig -2: Flow Chart

- **Advantages of Proposed system:**

1. Get Recommendation from overwhelming candidates within an acceptable time.
2. Avoid Reduplication Data's and improve performance the cluster Filtering.
3. Recommender Service System is Provide our knowledge provide Based from the users preferences.
4. The time of rating similarity computation between every pair of services will be greatly reduced.
5. The time complexity of Club CF can be divided into :
   - Offline cluster building.
   - Online collaborative filtering.

5. DEVELOPMENT METHODOLOGY

- **Naïve Bayes Algorithm:**

1. The Bayesian Classification represents a supervised learning method as well as a statistical method for classification. [4]
2. Assumes an underlying probabilistic model and it allows us to capture uncertainty about the model in a principled way by determining probabilities of the outcomes. [4]
3. Bayesian classification provides practical learning algorithms and prior knowledge and observed data can be combined. [4]
4. It calculates explicit probabilities for hypothesis and it is robust for noise in input data. It can solve diagnostic and predictive problems. [4]

- **Dictionary Generation:**

Count occurrence of all word in our whole data set and make a dictionary of some most frequent words.

- **Feature set Generation:**

All documents are represented as a feature vector over the space of dictionary words. For each document, keep track of dictionary words along with their number of occurrence in that document. Calculate Probability of occurrence of each label. Here label is negative and positive.

- **Training:**

In this phase we have to generate training data (words with probability of occurrence in positive/negative train data files). Calculate for each label. Calculate for each dictionary words and store the result (Here: label will be negative and positive). Now we have word and corresponding probability for each of the defined label.

6. IMPLEMENTATION

6.1 Twitter Module for Products:

Live tweets from the twitter are fetched directly and are stored in database. Fetching of tweets is based on unique tweet id or text assign to that product. After storing live tweets in the database, tweets are analyzed and categorized into 2 types: good and bad comments using Naïve-Bayes classifier. Based on that, every product is evaluated. If there are many good comments of a particular product, it is

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 05 | May -2017 www.irjet.net p-ISSN: 2395-0072

recommended to the user and if not, then that product is discarded and is not recommended to the customer.

```
Tomcat v7.0 Server at localhost [Apache Tomcat] C:\Program Files (x86)\Java\jre7\bin\javaw.exe (Apr 14, 2017 10:56:47 PM)
!!! Starting loop 3

Stop word remove from comment:- #4: Moto Play, 4th Gen (Black): Moto Play, 4th Gen (Black) (5825) Buy: Rs. 8,999.00
Stop word remove from comment:- #4: Moto Play, 4th Gen (Black): Moto Play, 4th Gen (Black) (5825)Buy: Rs. 8,999.00 R
Stop word remove from comment:- #4: Moto Play, 4th Gen (Black): Moto Play, 4th Gen (Black) (5825)Buy: Rs. 8,999.00 R
Stop word remove from comment:- Check Black Motorola MOTO 4 4th Gen Unlocked Cell Phone 16GB gift #Motorola https://
Stop word remove from comment:- Moto Play, 4th Gen (Black) https://t.co/qKrvx21HT8
Stop word remove from comment:- Moto Play (4th gen.) - Black - 16 GB - Unlocked https://t.co/cZBONI2jnJ https://t.co
Stop word remove from comment:- Check favorite Android Tech site. https://t.co/hqKt4qpfxg
Stop word remove from comment:- #deals Moto Play, 4th Gen (White) Free Power Bank selling cheaper INR 7950 today htt
Stop word remove from comment:- @AryanIndians Moto Play, 4th Gen (Black)?
Stop word remove from comment:- [Flat Rs. 2500 Off] Motorola Moto (4th Gen) - (2 GB RAM/16 GB) Rs. 10499 (Rs. 12999)
Stop word remove from comment:- BUY INR 599.00 Chevron Extremely Skinny Fashionable Rubberized Plastic Duvet Weights
```

Fig -3: Tweets are fetched

pid	result	tweet		tdate	tname	prate	nrate
Lenovo Vibe K5	normal	RT @Smartprix: Check ...	142 b...	Fri Apr 14 22:45:25 IST 2017	YJkFieLrdaUFXDk	0	0
Lenovo Vibe K5	normal	Touch screen issue le...	66 b...	Fri Apr 14 19:04:08 IST 2017	JaanMetsis	0	0
Lenovo Vibe K5	normal	Lenovo vibe K5 note i...	94 b...	Fri Apr 14 17:27:04 IST 2017	JaanMetsis	0	0
Lenovo Vibe K5	normal	RT @DigiTalkIndia: ? ...	142 b...	Fri Apr 14 17:16:48 IST 2017	herryks	0	0
Lenovo Vibe K5	normal	? 500 Discount on #Le...	130 b...	Fri Apr 14 17:15:51 IST 2017	DigiTalkIndia	0	0
Lenovo Vibe K5	positive	RT @srinudhulipalla: ...	144 b...	Fri Apr 14 16:45:49 IST 2017	Meenu804	1	0
Lenovo Vibe K5	normal	RT @buyakari: @santos...	96 b...	Fri Apr 14 16:45:44 IST 2017	Meenu804	0	0
Lenovo Vibe K5	normal	RT @santoshmnayar: @b...	142 b...	Fri Apr 14 16:45:42 IST 2017	Meenu804	0	0
Lenovo Vibe K5	normal	RT @buyakari: @Lenovo...	142 b...	Fri Apr 14 16:44:45 IST 2017	Meenu804	0	0
Lenovo Vibe K5	normal	I reviewed Lenovo-Vi...	111 b...	Fri Apr 14 14:34:25 IST 2017	Siddharth1738	0	0
Lenovo Vibe K5	normal	Heating issue in Leno...	55 b...	Fri Apr 14 14:12:07 IST 2017	JaanMetsis	0	0
Lenovo Vibe K5	normal	Lenovo Vibe K5 plus n...	65 b...	Fri Apr 14 13:04:36 IST 2017	JaanMetsis	0	0
Lenovo Vibe K5	normal	@lenovo Hi, my vibe k...	64 b...	Fri Apr 14 13:01:42 IST 2017	iamposhPARIS	0	0
Lenovo Vibe K5	normal	Lenovo Vibe K5 Note 3...	141 b...	Fri Apr 14 12:53:33 IST 2017	savemoney_india	0	0
Lenovo Vibe K5	normal	RT @Flipkart: Get a R...	141 b...	Thu Apr 13 12:32:38 IST 2017	rajat2571	0	0
Lenovo Vibe K5	normal	@lenovo Vibe K5+ is a...	142 b...	Fri Apr 14 10:21:42 IST 2017	buyakari	0	0

Fig -4: Tweets are stored in database

pid	postiverate	negativerate
Lenovo Vibe K5	75	25
Samsung On5 Pro	0	0
Lenovo Vibe K5 Snapdragon 616	0	0
Coolpad Mega 2.5D	100	0
Intex Aqua Ring	0	0
Samsung On7 Pro	0	0
Samsung Galaxy J7	100	0
Samsung Galaxy C9 Pro	100	0
Samsung Galaxy S7	100	0
SAMSUNG Galaxy J2 Ace	0	0
Apple iPhone 5s	0	0
Apple iPhone 6	100	0
Apple iPhone 7 Plus	80	20
Apple iPhone SE	100	0
Apple iPhone 7	100	0
Coolpad Note 5	100	0

Fig -5: Products are categorized as positive (good) or negative (bad)

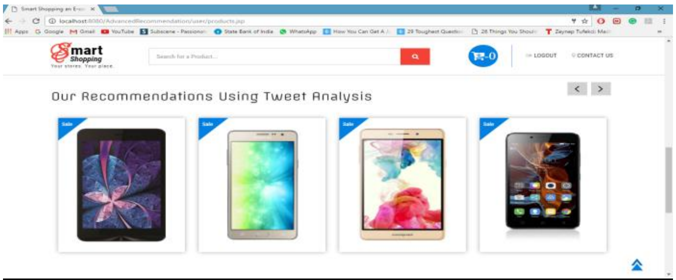

Fig -6: Recommendation based on Twitter

6.2 User History Module:

Transaction table maintains all the products purchased by all the users. In this module, hash function is used to know liking of a particular user based on his/her purchase history. Input for this hash function is retrieved from transaction table. Mapping of hash function is done to find the proper products to recommend. To increase the efficiency, iterative sorting of this hash function is used and the result obtained after this step is recommended to the user/customer.

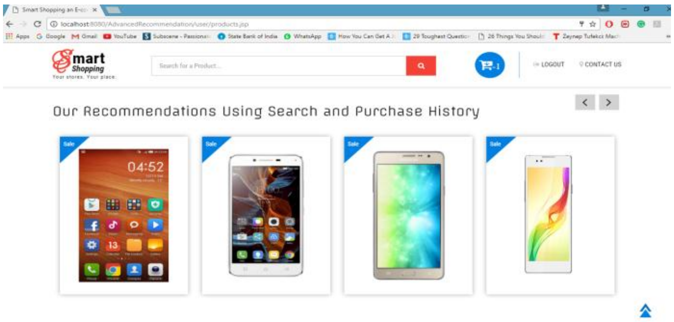

Fig -7: Recommendation based on User Purchase History

6.3 Rating Module:

Various users rate some products. These ratings are stored in database. Ratings for a particular product are taken and average of all these ratings is calculated. This average is stored in rate table of database. Above process is done for every product that is rated and for recommendation, products are selected based on descending order of average ratings.

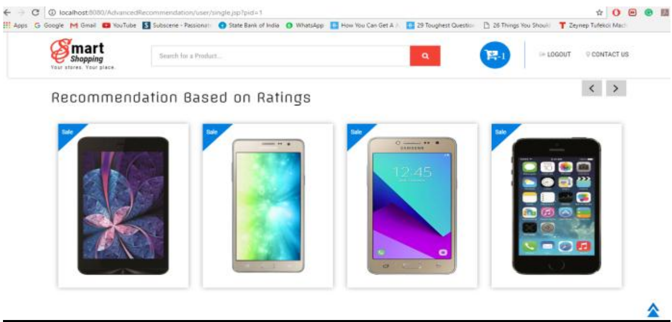

Fig -8: Recommendation based on Ratings

7. ANALYSIS

Table -1: Result Analysis

Existing System	Implemented System
(1).In Existing System, decision making is time consuming.	(1).In implemented system, decision making is spontaneous compared to Existing System.
(2).The cluster analysis gathers users with	(2).The cluster analysis gathers users with

similar characteristics according to the web visiting message data only.	similar characteristic based on ratings, popularity, social media library along with web visiting message data.
(3).The performance of Time to search and data clustering management is poor.	(3).The performance of Time to search and data clustering management has improved significantly.
(4).In many of the Existing systems, they use only a specific type of recommendation technique so the recommendation is poor.	(4).In the Implemented system, many techniques have been combined together so the recommendations are accurate.
(5).Using only one specific technique for recommendation results in reduplication of data.	(5).By using various techniques, implemented system avoids reduplication of data and improves performance of the Collaborative Filtering.
(6).In the Existing System, the time of rating similarity computation between every pair of services is quite high.	(6).In the Implemented System, the time of rating similarity computation between every pair of services will be greatly reduced.

8. CONCLUSION

We present a collaborative filtering approach for applications relevant to service recommendation. Before applying CF technique, services are merged into some clusters via a data mining algorithm. Within the same cluster, rating similarities between services are computed. As the number of services in a cluster is much less than that of in the whole system, CF costs less online computation time. Moreover, as the ratings of services in the same cluster are more relevant with each other than with the ones in other clusters, prediction based on the ratings of the services in the same cluster will be more accurate than based on the ratings of all similar or dissimilar services in all clusters. These two advantageous of CF have been verified by experiments on real-world data set.

9. FUTURE SCOPE

Future research can be done in two areas. First, in the respect of service similarity, semantic analysis may be performed on the description text of service. In this way, more semantic-similar services may be clustered together, which will increase the coverage of recommendations. Second, with respect to users, mining their implicit interests from usage records or reviews may be a complement to the explicit interests (ratings). By this means, recommendations can be generated even if there are only few ratings. This will solve the inadequacy problem to some extent.

REFERENCES

[1] A. Bellogín, I. Cantador, F. Díez, P. Castells, and E. Chavarriaga, ``An empirical comparison of social, collaborative filtering and hybrid recommenders," ACM Trans. Intell. Syst. Technol.,vol. 4, no. 1, pp. 1_37, Jan. 2013.

[2] T. C. Havens, J. C. Bezdek, C. Leckie, L. O. Hall, and M. Palaniswami, "Fuzzy c-means algorithms for very large data" In Trans. Fuzzy Syst., 2012 IEEE International Conference on, pp. 1130-1146. IEEE, 2012.

[3] W. Zeng, M. S. Shang, Q. M. Zhang, L. Lü, and T. Zhou, ``Can dissimilar users contribute to accuracy and diversity of personalized recommendation?," Int. J. Modern Phys. C, vol. 21, no.10, pp. 1217_1227, Jun. 2010.

[4] Liu, Bingwei, Erik Blasch, Yu Chen, Dan Shen, and Genshe Chen. "Scalable sentiment classification for big data analysis using naïve bayes classifier" In Big Data, 2013 IEEE International Conference on, pp. 99-104. IEEE, 2013.