Efficient Recommendation System based on Keyword for Smart-City with Sentiment Analysis using Hadoop

Ms. Sneha G. Potdar, Prof. N. S. More

Dept. of Information Technology, SKNCOE,
Pune, India.

Abstract *— In today's life recommender system plays the important role while user is on social networking site, online shopping website. For presenting the personalized recommendation according to the new user's demand is big challenge. The previous recommender system produces the recommendation without checking the personalized interest of user, due to that they couldn't meet their proper requirement. In this paper, the recommender system works very efficiently using Hadoop by computing the reviews based on removing stop-words, stemming algorithm and sentiment analysis algorithm. Therefore, in the Smart-City application, the user getting accurate recommendation by computing reviews in positive, negative and neutral way according to new user's keywords or sentences with the related services. Therefore, this paper approaches speed-up the performance over large amount of dataset using Map-Reduce. The proposed system puts forward an idea of aspect parsing technique and sentiment analysis technique which is forwarded with precision and recall measurement for an effective and accurate recommendation over reviews.*

Keyword — *Accuracy, Efficient, Keyword, Recommender System, Smart-City, Sentiment Analysis, Hadoop, Map-Reduce.*

I. INTRODUCTION

The increasing no. of web services and increasing no. of internet users are growing rapidly. The data is collected from various sources like sensors, machine data, the data from the social networking, etc. The different varieties of evidence, the large volume of data and the velocity of data can cause the growth of data day by day. The Hadoop can process large volume of data very quickly. The recommendation system is the adaptive suggestion system to the user according to their interest. Here most of the recommender system can recommend the services to the user without analysing the previous reviews and current ratings. So the new user fails to meet the personalized service. The Keyword Aware Recommendation System (KASR) presenting the personalized recommendation based on the keyword[1]. Keyword service recommendation technique is presenting a personalized service recommendation list and recommending the most appropriate service to the users. Keywords are used to indicate user's preferences, and a User-based Collaborative Filtering algorithm is adopted to generate appropriate recommendations. The KASR aims at calculating a personalized rating for each candidate service for a user, and then presenting a personalized service recommendation list and recommending the most appropriate services to him/her. The Smart-City recommendation system is having the list of the Smart Cities along with their services. The user can select the city and choose the service from the given list. This preference of the user is nothing but the keyword. Here keywords in the form of single keyword, multi-keyword or the sentence. This keyword can computes with previous users' similar preference from the database by using parsing and stemming techniques. If the previous reviews present in database relate to that keyword, then reviews are categorizing by using the sentiment analysis. The Map-Reduce is using for providing the most accurate and efficient recommendation to the user by computing the large set of reviews.

The remaining paper is organized as follows: Section II describes the related work of various recommendation systems. Section III presents the proposed methodology for Smart-City Recommendation. In the section, IV states results and discussion and section, V presents conclusion and future work.

II. RELATED WORK

Badrul Sarwar et. al. [2], here author concerns for the growth of business, the E- Commerce is using the group recommendation technique for providing better service to the user. Here they make the cluster of the user depending on their same types of reviews. The clustering of the population will divide the reviews of the user, and because of that, it gets affected by the accuracy of the service. Yan-Ying Chen et. Al. [3], here author works on the travelling recommendation it gives the user personal level recommendations for their trip. They keep track of the photos which they were clicked from various locations and depend on that they can give the recommendation to the user on the personal level.

They didn't consider the group level preferences for computation, and this is the limitation of the traveling system. Xin Cao et. al. [4], While searching the route from source to destination the user must have to give two specific keywords as source keyword and destination keyword. In the Google map by using these two keywords, it will recommend you the different route options. The user will choose the convenient route for his/her journey. J. Amaithi Singam et. al. [5], In the optimal keyword search, the issues occurring when service recommender system implements in large data sets. It splits the services to the users and mainly focused keywords from the user's preferences. For generating keyword recommenders from the previous user preferences here using Hybrid Filter Algorithm. The result is shown here on Real-World datasets and reduces the processing time from large datasets. T.N. Chiranjeevi et. al. [6], PRS solving the challenges of the existing system like it provides the same result to the user's based on the evaluation and ranking or rating service. By using the collaborating filtering and Porter stemmer algorithm, it gives the suitable recommendation to the user. Here Personal Recommendation System is considered user's preference and necessity. This is using the Hashmap technique for faster keyword search for selecting correct reviews, and indexing method also used here in eliminating the articles like a, an, the, etc. X. Yang et. al. [7], The Bayesian-inference gives the recommendation for the social network. Here the user is get embedded with each other in the network. It is providing the accurate and personalized recommendation to the user. Each and everything is based on the network so the human being is facing difficulty for facing such interface. Xiaoyuan Su et. al. [8], In the survey of collaborative filtering techniques, there having the study of various techniques. There are three types of CF techniques such as Memory based technique, Model-Based Technique, Hybrid Based Technique. Each CF technique works in a different way. Yamini Nikam et. al. [9], In the survey on Service Recommendation Technique there are various recommender techniques such as Content-based, Knowledge Based, Social Network Based, Context Awareness based, Group Based, and Keyword Aware Service Based Recommendation. These recommendations are based on any one of the collaborative filtering technique. It is improving the scalability and the accuracy of the traditional systems. Khushboo R. Shrote et. al. [10], [11] The sentimental analysis is known as the opinion mining analysis. It is using for analyse the users opining. This analysis is used for computing the positive and negative reviews from the user. And based on the analysis it provides the proper result. This survey is only about for how to apply sentimental analysis on the Keyword aware Service Recommendation for getting the appropriate result[12]. Rita Guimaraes et. al. [13], It presents a recommendation system based on sentiment analysis of sentences extracted from Social Networks considering an algorithm which depends on the adverb found in the sentence. Depending on that adverbs the algorithm can performing the analysis. Here the recommendation system has low complexity and presents low perceived impacts on the analysis of energy consumption according to benchmark software. Kazuyoshi Yoshi et. al. [14], A hybrid music recommender system that ranks musical pieces by comprehensively considering collaborative and content-based data that is rating scores derived from users and acoustic features derived from audio signals. This mechanism is using his or her own preferences to select the music according to their choice. System accurately recommended pieces including nonrated ones from a wide variety of artists and maintained a high degree of accuracy even when new users and rating scores were added. Faustino Sanchez et. al. [15], It presents a Recommender System for Sport Videos. Here the new recommendation method transparent to the user, who only has to consume videos as he or she would do in any video distribution platform. The system takes into account how the preferences of users change over time and based on that the recommendation was provided. This system is integrated on the client-side, and it avoids a lot of computational problem and congestion. Dr. M. Durairaj et. al. [16], In the news recommendation system it is using the content, knowledge and collaborative based techniques. The contents of the particular articles that a user has read in the past by analysing that with the help of content based approach. By using the collaborative filtering technique, the system makes the recommendation to the others, using known preferences of a group of the users. Shunmei Meng et. al. [1], here author gives details about Keyword Aware Service Recommendation, which is using for different types of domain thesaurus. And in KASR it will recommend the service by their previous and current reviews of the user for Hotel Reservation. Here keywords are used to indicate the user's preferences. The previous preferences of the user will be considered here for giving the appropriate service recommendation. It is using in Travelling recommendation, Hotel Reservation, Online Shopping, etc., and because of that, it is in the traditional way.

III. PROPOSED WORK

This section describes the detailed implementation of tool. Here Fig 1. shows how the recommendation System works for Smart-City Application accurately and effectively.

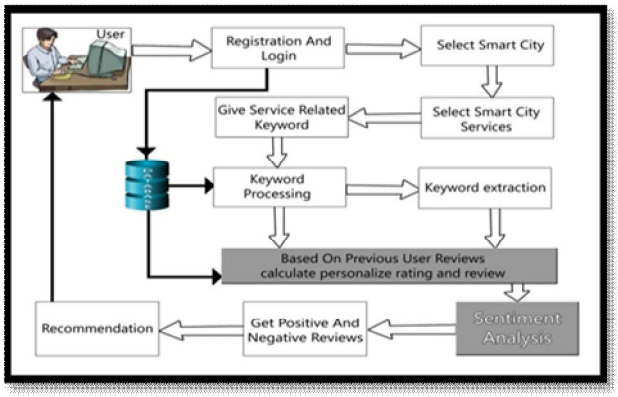

Fig. 1: System Framework for Smart-City Recommendation

The following are the steps for recommendation System based on Keyword.

A. Capturing the Users Preferences

a) Preferences by Active User

The active user will give his/her preference by selecting the keyword from a Candidate-Keyword list. The active user should also select the keyword which having importance degree of the keyword by checking its review and rating.

b) Preferences by Previous User

The preferences of the previous user can be extracted for a candidate service according to the candidate keyword list and domain thesaurus.

Here, these two algorithms are useful after capturing the users preferences. One is used for removing the stop words while another is gives the main root word by stemming the word.

TABLE I
ALGORITHM FOR STOP-WORD

Input: PPK and APK keyword. APK={apk1,apk2,…….apkn} PPK={ppk1,ppk2,ppk3…….ppkn}
Step 1: Accept the stop word from the database Step 2: Stopwords={st1,st2,st3……,stn} Step 3: Compare the APK word with stopwords Step 4: If APK word equal to stopword Step 5: Remove apki Step 6: Compare the PPK word with stopwords Step 7: If PPK word equal to stopword Step 8: Remove ppki Step 9: Return without_stop_word_keyword_set_of_APK Step 10: Return without_stop_word_keyword_set_of_PPK

The above TABLE. I shows Stop-Word Algorithm, it is using for removing the stopwords from the selected reviews. It removes the words like a, an, the, of…etc.

TABLE II
ALGORITHM FOR STEMMING WORD

Input: without_stop_word_keyword_set_of_APK={k1,k2,k3…….kn} without_stop_word_keyword_set_of_PPK={k1,k2,k3,……kn}
Step 1: Select ending characters of word. Step 2: Examine the final letter. Step 3: Check applicability of the rule of stemming word. Step 4: The ending characters are not matched by the stemming rule word then words are not replaced by the natural language. Step 5: Apply rule. Step 6: Delete the end character of keyword and replaces by the satisfied word you want. Step 7: Return StemWord

Here, the Stemming word algorithm is using for finding the root word for the particular word. It removes the ending character from the word and gives the natural word by stemming. For example, the word beautiful is having its root word beauty.

B. Similarity Computation

It is the second step to identify the reviews of the previous user who have similar choices to an active user by finding neighborhoods of the active user based on the similarity of the preferences.

By using two computation methods giving the recommendation to the user can be performed. The methods are as follows: a) Approximate Similarity Computation Method and b) Exact Similarity Computation Method.

a. Approximate Similarity Computation

Here comparing the similarity and diversity of sample sets, Jaccard Coefficients[1], is applied in the approximate similarity computation.

$$sim(APK, PPK) = Jaccard(APK, PPK) = \frac{|APK \cap PPK|}{|APK \cup PPK|}$$

Where

APK is the preference keyword set of the active user,

PPK is the preference keyword set of a previous user.

b. Exact Similarity Computation

A cosine based approach is applied in the exact similarity computation, and it is similar to vector space model in information retrieval.

$$w_{pk_i} = TF \times IDF = \frac{N_{pk_i}}{\sum_g N_{pk_i}} \times \log \frac{|R'|}{|r' : pk_i \in r'|}$$

Now, based on this weight it is possible to calculate the Exact Similarity computation of keyword.

TF- The 'Term Frequency' is no. of occurrences of the keyword pk_i in all the keyword sets of the reviews commented by the same user and g is the no. of preferences of the keywords.

IDF- The Inverse Document Frequency is obtained by dividing the number of all reviews by the no. of reviews containing the keyword pk_i.

C. Sentiment Analysis

In Similarity computation, it considered the previous and active user's preferences for computing appropriate recommendation, likewise in sentimental-analysis, it categorizes the positive and negative reviews of the previous users. Because of sentimental analysis the system works in more efficient way[12]. In proposed system by using various parameters like systems speed, efficiency, scalability, accuracy, etc. will be computing on live data for better recommendation performance. The following TABLE. III. Shows the algorithm for sentiment analysis is given below in, which states steps of sentiment analysis.

TABLE III
ALGORITHM FOR SENTIMENT ANALYSIS

Input: previous users review R={r1,r2,r3,…...rn}
Output: List of requested sentiment
Step 1: If R>0 then
Step 2: For each I to R
Step 3: Using the parser to split the review R into positive review
Step 4: ServiceRelatedKeyword Sk(sk1, sk2, sk3...skn)
Step 5: for each(R in Review)
PositiveSentiment = true
Sum=Sum+PositiveSentiment
//Store sum in DB
Step 6: For each I to R
Step 7: Using the parser to split the review R into negative review
Step 8: ServiceRelatedKeyword Sk(sk1, sk2, sk3...skn)
Step 9: For each(R in Review)
if (R == keyword)
NegativeSentiment= true
Sum=Sum+NegativeSentiment
//Store sum in DB
Based on Sum
Step 10: Display the proper recommendation related to particular service
Step 11: Else
Step 12: No review available
Step 13: Return Positive review

Following are the Notations used to represent sentiment analysis mathematically:

Where,

R- Review
Ski- Service-Related keyword
WA- splitting review store in word array
Dir- Directory of words
PSA- Positive Sentiment Analysis
NSA- Negative Sentiment Analysis
Sum_P- sum of positive Sentiment Analysis
Sum_N- sum of negative Sentiment Analysis

Give the review R as input to proposed system,

R = {r1,r2,r3,…....,rn}

On the basis of 'R', it should not Null, $R \neq 0$

The Review checks upto I..Rn is positive sentiment

$$SKi = \sum_{i=0}^{rn} split(R)$$

Declare *WA* variable as null,

$$WA < -\emptyset$$

$$WA < -SKi$$

It takes system define Dictionary as full word array,

$$Dir < -LuceneWordNet$$

Now checks the sentiment of each WA words with Dir,
If words evaluated by equation 2 and Dir matches, then word is positive

$$PSA = \sum_{i=0}^{WA} (WA \in Dir) in (length_{WA} > i > 0)$$

If words evaluated by equation 2 and Dir matches, then word is negative

$$NSA = \sum_{i=0}^{WA} (WA \in Dir) in (length_{WA} > i > 0)$$

Now sum of the all positive and negative sentiment of Service related keyword,
Sum_P=Sum_P + PSA Sum_N=Sum_N + NSA
Now based on the Sum_P and Sum_N of the active user and previous users review further filtering will be conducted.
PSA=Sum_P > Sum_N

NSA=Sum_N >Sum_P

If Sum_P=Sum_N=0, then the active user is the first user who is giving the review to that keyword.

D. Generate Recommendation

The recommendation would be generated to the user by calculating the similarity of the active user and previous user and previous user. Here calculating the personalized rating based on the similarity computation and by using sentiment analysis. Finally, a personalized service

recommendation list will be presented to the user and the service(s) with the highest rating(s) will be recommended.

Here following Fig. 2. is related to the simple domain thesaurus of Smart City System, in that set the Candidate-Keywords which is denoted in the Rectangle and the keywords which in the Oval is known as the Related-Keywords.

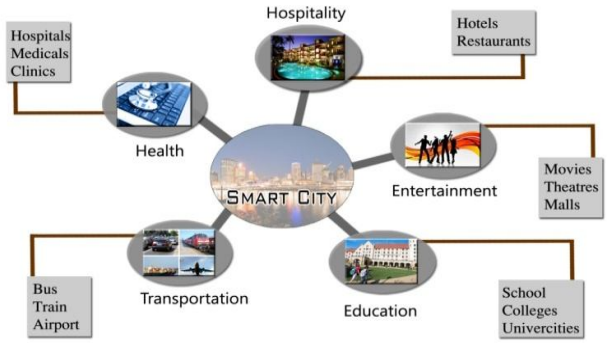

Fig. 2 Block Diagram of Smart City and its services

i. Candidate and Related Keyword List

The Candidate-Keyword is the list of the Keywords about the user's preferences, and multi-criteria of the candidate services, which can be denoted as K={k1,k2,….,kn}, n is the no of keywords in the Keyword-Candidate list.

TABLE IV
CANDIDATE AND RELATED KEYWORDS

Sr. No	Candidate Keyword	Related Keyword
1	Entertainment	Movies Theater Malls
2	Education	School High-Schools Colleges University
3	Hospitality	Hotels Restaurants
4	Transportation	Bus Taxi Train
5	Health	Hospital Medical Clinic

These are the Candidate-Keywords with its Related-Keywords according to the smart city System.

IV. RESULTS AND DISCUSSION

Here, some of the experimental evaluations are performed to show the efficiency and effective work of the system. These evaluations are performed on Windows based java machine using Eclipse.

The Smart City Application introduces, giving the best recommendation of the current or new user to find the services. For that purpose in the proposed system here using stemming, stop words, and sentiment analysis algorithm to parse the review with the help of Map-Reduce.

The proposed system working on the datasets of Hotel in that it added 50 services and more than 1000 review. The Fig. 3. showing ratio of average rating for services, Fig. 4. showing the pie-Chart for review after Sentiment Analysis and Fig. 5. Showing the time graph for running services on Hadoop. And the effectiveness of the system is measured in Precision and Recall which shows in Fig. 6.

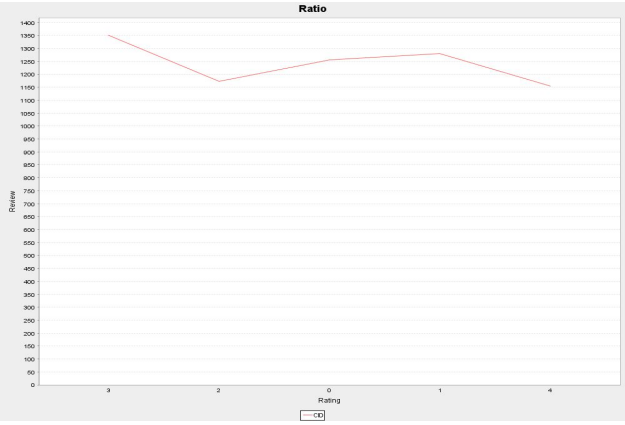

Fig. 3 Ratio of Average Rating

The above Fig. 3. gives the details about an average rating of an overall dataset according to their consumer rating over no. of reviews.

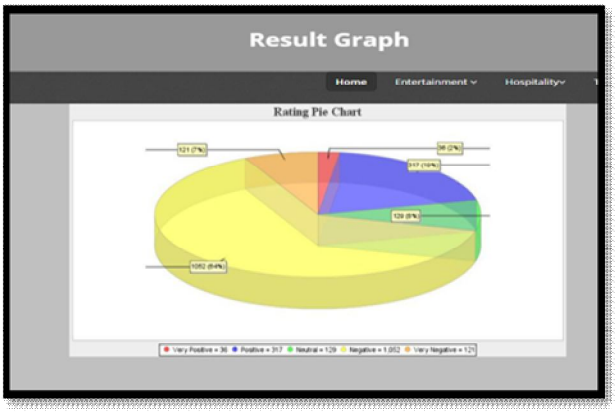

Fig. 4 Rating Pie Chart for Review after Sentiment Analysis

The above Fig 4. in that the pie-chart is generated after the sentiment analysis performed according to the active user. It gives details regarding the reviews that how many reviews are under which category of sentiment. It is categorized within five sentiments that very positive, Positive, Neutral, Negative, Very Negative.

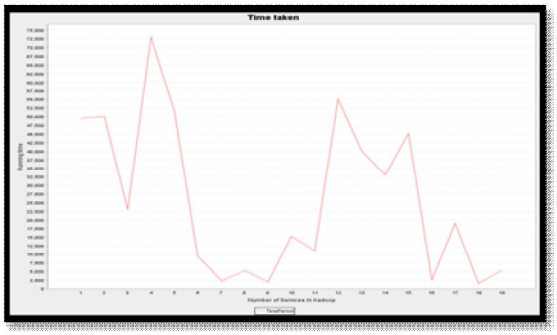

Fig.5 Timing Graph for Running Services

Here, the Fig. 5 gives the timing graph for running services on Hadoop. The running time period is measured in the millisecond. Hence it is proved that Hadoop can perform faster computation over large datasets.

Recall and Precision

This calculation is based on true positive reviews and false positive reviews. True negative reviews can be calculated by the sentiment analysis rating with the help of the aspect that calculated by the aspect parser.

The formulae for Precision and Recall is as follows:

TABLE V
RECALL AND PRECISION FORMULAE

int TP= veryPositive; int FP= neural-negative;	int TP= veryPositive; int FN=VeryNegative;
Precision = TP/TP+FP;	Recall=TP/TP+FN;

The Recall is defined as the number of relevant reviews retrieved by a search divided by the total number of existing reviews documents. And the Precision is defined as the number of relevant reviews retrieved by a search divided by the total number of reviews retrieved by that search.

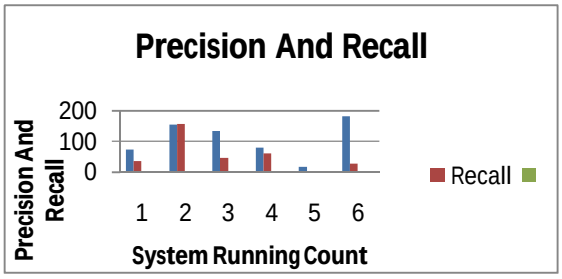

Fig.6 Precision and Recall Graph for Keyword

Thus in Fig. 6, the Precision and Recall is measured for the smart-city recommendation based on the sentiment analysis rating. The higher rating is supposed here as Very Positive rating, and the lower rating within the same aspect parsing table is considered as a Very Negative rating. Based on the calculation formulae for Precision and Recall the effectiveness and accuracy of the system shown in above graph.

V. CONCLUSIONS AND FUTURE WORK

The proposed system used the sentimental analysis for the user's accurate review predictions according to the keywords; it has been found that the system allows sentimental analysis for dividing the reviews into the positive, negative category which improved its efficiency by using categorization of keywords, and the measurement is done by using recall and precision. The future the recommendation system may compute the reviews by converting the video and audio data into text for better recommendation using sentiment analysis.

ACKNOWLEDGMENT

I hereby take this opportunity to express my heartfelt gratitude towards the people whose help was very useful for completion of my research work on the topic of "Efficient Recommendation System based on Keyword for Smart-City with Sentiment Analysis using Hadoop". It is my privilege to express sincerest regards to the project guide Prof. N. S. More, for his valuable guidance. I deeply express my sincere thanks to our all faculty members of "SKN College of Engineering" for their support. I would like to thank my parents as well as my friends for their encouragement and supports.

.

REFERENCES

[1] Shunmei Meng, Wanchun Dou, Xuyun Zhang, and Jinjun Chen, Senior Member, IEEE "KASR:A Keyword-AwareService Recommendation Method on MapReduce for Big Data Applications", IEEE Transactions on Parallel and Distributed Systems, VOL. 25, NO. 12, DECEMBER 2014.

[2] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl, "Item Based Collaborative Filtering Recommendation Algorithms" *ACM, 10*, May 15, 2001.

[3] Y. Chen, A. Cheng and W. Hsu, "Travel Recommendation by Mining People Attributes and Travel Group Types From Community-Contributed Photos". IEEE Transactions on Multimedia, Vol. 25, No.6, pp. 1283-1295, 2012.

[4] Xin Cao, Lisi Chen, Gao Cong, Jihong Guan, Nhan-Tue Phan, Xiaokui Xiao, "KORS: Keyword-aware Optimal Route Search System", ICDE Conference, pages 1340-1343,IEEE 2013.

[5] J. Amaithi Singam and S. Srinivasan, "Optimal Keyword Search For Recommender System In Big Data Application", ARPN Journal of Engineering and Applied Sciences, VOL. 10, NO. 7, APRIL 2015.

[6] T.N. Chiranjeevi, R.H. Vishwanath , *"PRS: personnel recommendation system for huge data analysis using porter stemmer "*, ICTACT JOURNAL ON SOFT COMPUTING, VOLUME: 06, ISSUE: 03,April 2016.

[7] X. Yang, Y. Guo, Y. Liu, "Bayesian-inference based recommendation in online social networks", IEEE Transactions on Parallel and Distributed Systems, Vol. 24, No. 4, pp. 642-651, 2013.

[8] Xiaoyuan Su and Taghi M. Khoshgoftaar, *"A Survey of Collaborative Filtering Techniques"* Hindawi Publishing Corporation, Advances in Artificial Intelligence, Article ID 421425, 19 pages doi:10.1155/2009/421425, 2009.

[9] Yamini Nikam, M. B. Vaidya,*"Survey On Service Recommendations Techniques"* IJARIIE, Vol-2, Issue-1,2016.

[10] Khushboo R. Shrote, Prof A.V.Deorankar, "A Survey on Sentiment Analysis Based Keyword Aware Service Recommendation for Big Data", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 10, October 2015.

[11] Mr. T. Rajasekaran, P.Barathkumar, M.Guruprasath, S.Santhosh, "Keyword Conscious Provision Endorsement Method ", IJLTEMAS, Volume IV, Issue III, March 2015.

[12] Khushboo R. Shrote, Prof A.V.Deorankar, "Sentiment Analysis Based Hotel Recommendation System", Paripex - Indian Journal Of Research, Vol. 5, Issue 4 , April 2016.

[13] Rita Guimaraes, Demóstenes Z. Rodriguez, Renata L. Rosa, Graça Bressan, "Recommendation System using Sentiment Analysis considering the Polarity of the Adverb", IEEE International Symposium on Consumer Electronics, 2016.

[14] Kazuyoshi Yoshii, *Student Member*, *IEEE*, Masataka Goto, Kazunori Komatani, Tetsuya Ogata, *Member*, *IEEE*, and Hiroshi G. Okuno, *Senior Member*, *IEEE*, "An Efficient Hybrid Music Recommender System Using an Incrementally Trainable Probabilistic Generative Model", IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 16, NO. 2, FEBRUARY 2008.

[15] Faustino Sanchez, María Alduan, Federico Alvarez, Member, IEEE, Jose Manuel Menendez, Member, IEEE, and Orlando Baez, "Recommender System for Sport Videos Based on User Audiovisual Consumption", IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 14, NO. 6, DECEMBER 2012.

[16] Dr. M. Durairaj, K. Muthu Kumar, "New Recommendation Systems Using Web Mining:A Study", IJETT, Vol 12, No. 6, Jun 2014.