ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

Decision Making System for User Behaviour Analysis using Prediction Algorithm

Hemlata Kardas, Sulochana Sonkamble

Student, Department of Computer Science, Rajarshi Shahu College of Engineering & Research, JSPM Narhe, Technical Campus, Pune university, Pune, India.

HOD, Department of Computer Science, Rajarshi Shahu College of Engineering & Research, JSPM Narhe, Technical Campus,Pune university, Pune, India.

ABSTRACT: Various organizations need Decision Making System for enhancing the service providing ability towards user. So, the user behaviour analysis is required for decision making system to deal with understanding requirement of users. Also to provide efficient services through number of domain like product based system, reservation systems and mobile system etc. The analysis includes extraction of data related to various domains like ratings, rankings, reviews, post or any other information which is related to users feedback and product or system feature. In this paper Hotel Reservation System is used to provide personalized list of hotels to user from review processing by mining data in terms of keywords extraction. The keywords are nothing but user preferences. Then prediction algorithm called Keyword Aware Service Recommendation method is used to predict the behaviour of user based on occurrence of keywords in reviews. This is user-based collaborative filtering algorithm used for prediction which determines the similarity of set of keywords referred by current user and previous user. As the review analysis has big amount of data, it is processed using Hadoop environment and Map Reduce parallel programming framework to enhance the scalability and efficiency of KASR method. Finally experimental results are achieved based on hotel reviews dataset used for demonstration which will improve KASR for accurate and scalable user behaviour system.

KEYWORDS: user preferences, keywords, recommendation, big data, Map Reduce.

I. INTRODUCTION

The Decision making system is need of various organizations which has the predictive analysis mechanism for decision making activities of various organizations to improve their performance about the service they provide. Web usage mining also gives knowledge about user's behaviour towards service by extraction of interesting patterns for analysis. Now-a-days there is huge amount of data is seen over internet which is to be processed to provide efficient service to user and also to analyze the user's requirement. IT industry has challenge to manage this big data which is key basis for productivity growth, innovation and customer need satisfaction. There are several service recommender systems available for big data application. But research has been done on developing new approaches for the same. One of the recommender systems like Hotel Reservation System which analyses user's preferences by collecting the reviews posted to get personalized list of hotels as per their requirement. This system is based on user based collaborative filtering algorithm where it recommend services to the user that users with similar tastes preferred in the past. It supports Hotel Reservation System through keyword based algorithm designed for it which extracts set of keywords commonly used while searching good hotels by user. Here keywords are nothing but users preferences captured form previous users post and provided to current user for efficient processing. Then it performs processing on keyword set by comparing with users preferences based on Keyword Aware Service Recommender method. This Method provides personalized list of hotels to users. So as to enhance efficiency and scalability KASR algorithm is put on Hadoop and MapReduce framework for parallel processing [1].

There are various methodologies are seen for variety of domains to analyze users patterns for understanding their requirements and provide them efficient services. Some of the author did retrieval of product aspects and opinion through customer reviews generating product feature using unigram language model. Then they did the mapping of opinion to product feature [2].One of the method focused on ranking product features by designing product ranking framework to identify aspects and define semantic classifier with probabilistic model for comparing the opinion with

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

overall opinion [3]. Some author defined the cross-domain thesaurus by using semi supervised data which is used to generate domain thesaurus randomly. [4]. Some authors have estimated sales impact performance based on reviews by computing helpfulness value using votes given by user for that product [5]. Specifically, some authors worked on generating mashup tool for processing user behavior using interaction, confidence, diagnostic, intention to analyze decision quality [6]. Some authors did multi criteria user modeling to improve the quality of service [7] and online review processing for movie domain to understand public sentiment and business intelligent [9].

II. RELATED WORK

Lisette García-Moya et al. focused on identification of aspects or feature of product from customer's reviews about product features, semantic classification from opinion of customer and aspect ranking is identifying relevance of aspect and opinion. This system considers stochastic mappings between words to estimate a unigram language model of product features. It determines the probabilistic model for mapping opinion to product feature by retrieving words from reviews based on co-occurrence vale and refining them. Finally evaluation of retrieval is done using HITS method [2]. Zheng-Jun Zha et al. focused on extracting product feature and ranking them using probabilistic aspect ranking based on overall opinion rating by weighted aggregation of opinion on aspects. This model uses importance of aspect while aggregating words. This system uses *sim* function for product aspect identification for finding occurrence of words from noun and phrases. Also uses language model for scoring the aspect and semantic classifier which parses the opinion of user and generate set of aspect using lexicon method [3]. Danushka Bollegala et al. focused on automatic classification of semantic for various applications for opinion mining. This system constructs sentiment sensitive distributional thesaurus by labeling source data and unlabeled data in target domain for cross domain sentiment classification. It also expands the feature vector for enhancing the domain thesaurus. This system uses semi supervised method for classification of domain. The relatedness of reviews is computed using POS tagging and unigram, bigram model [4] [10] [11]. Anindya Ghose et al. focused on quality determination of reviews by text mining. Random Forest-based classifiers method is used for predicting reviews based on the decision process system which uses "*helpfulness*" value. Reviews basically have objective and subjective features used for computing *helpfulness*. Here objective features are nothing but characteristics or description of product by merchant and subjective features means reviews or personal opinion by customer. It then finds the probability of occurrence of text words in subjective and objective features which are used to identify salesrank based on reviews. The *helpfulness* value is computed by ratio of votes to total votes received for product [5]. Brandon A. Beemer et al. focused on dynamic interaction of user for determining the quality of service of product by revising and revisiting the inputs for changing decisions. The relationships between dynamic interactions, diagnostic, confidence, and intention are analyzed here using mashup tool. Post hoc analysis of decision quality suggests that increased levels of dynamic interaction also improve the overall quality of the decision made [6]. Kleanthi Lakiotaki et al. defined set of phases for user modeling is with first phase data acquisition gathers data in terms of numerical rating and ranking forming data matrix. Second phase multi criteria user modeling does aggregation of multi criteria data using UTA method providing weight vector with user modeling. Third phase is clustering of data object and their relationship with each other using k-means algorithm. There set of clusters are fed to fourth phase called recommendation where it applies sim function on clusters to identify the similarity of users behavior using collaborative filtering algorithm. Here multi criteria similarity computation is based on multiple matrices like statistical or classification accuracy [7].

Above all the methodologies worked on various different prediction models for user behavior analysis. But KASR method works with collaborative filtering prediction algorithm with Hadoop and Map Reduce technique which innovative one because it deals with Big Data environment for efficient result while processing through it.

III. PROPOSED ALGORITHM

A. *Design Considerations:*

- Initially required to collect reviews for capturing previous users preferences through dataset like posted reviews of some hotels.
- Mining keywords through text mining method like removing stop word and stemming words.
- Provide interface for current users to allow them to choose required keywords of hotels like Food, Service, Location, Transportation and etc.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

- Apply KASR method which has two different algorithm based on Similarity Computation by Jaccard Coefficient using Approximate Similarity Computation (ASC) algorithm and cosine similarity using Exact Similarity Computation (ESC) algorithm.
- Further ASC and ESC will perform MapReduce Computation on set similar of keywords extracted as above using Mapper and Reducer class.
- By forming combination of key value pair of keywords and occurrence value, KASR outputs personalized set of hotels to user using MapReduce method.
- But there is chance to improve the results by analyzing positive and negative preferences of user while processing datset.

B. *Description of the Proposed Algorithm:*

Aim of proposed method is to deal with ASC and ESC algorithm to execute on Mapreduce framework. And trying to distinguish positive and negative preferences by forming combinations of keywords related to sentiments like good, bad, dirty, dark, dislike etc. Then provide this set of keywords either during process of stop words removal or during Mapreduce processing and finally to provide personalized list of hotels to users as recommended by them.

Step 1: Mining of review data for obtaining Keyword:
This process involves Stemmer method for stemming words which transforms root word to required single word by removing stems like ing, full, less, ed, es, ness, tive etc and Removal of stop words involves providing stop words list with dataset for comparing stop words form dataset and then remove it.

Step 2: Perform processing using KASR method using two algorithms as follows:

a. Approximate Similarity Computation (ASC) algorithm: This ASC algorithm is based on processing of comparison of keyword set of current user (CU) and previous user (PU) *Jaccard coefficient*, This Jaccard coefficient is measurement of asymmetric information on binary (and non-binary) variables, and it is useful when negative values give no information. Basically it is used to compare similarity and diversity of keyword set.

$$sim(CU, PU) = jaccard(CU, PU) \frac{|CU \cap PU|}{|CU \cup PU|} \dots\dots\dots\dots\dots\dots\dots eq.(1)$$

b. Exact Similarity Computation (ESC) algorithm: This ESC algorithm is based on capturing exact similar set of keywords form CU and PU keyword set by comparing with *cosine similarity function.* Here weight computation of keywords by assigning importance to keywords based on words commonly used by user and TF/IDF computation for frequency of occurrences of keyword set from current user and previous user preferences.

$$sim(CU, PU) = \cos(CU, PU) = \frac{CU.PU}{||CU|| \times ||PU||} \dots\dots\dots\dots\dots\dots eq.(2)$$

Step 3: Personalized recommendation list generation: The similarity of current and previous user preferences is determined by comparing *sim (CU, PU)* value with threshold ∂ and further filtering is done. Once the most similar keyword list is obtained personalized list is computed. This uses weighted average method.

Step 4: Map Reduce Implementation:

- Once formulation has been done it is fed to map () and reduce () function along with dataset. In this system it starts with executing Approximate Similarity Computation algorithm (ASC) using Hadoop.

1. In first phase, Hadoop dose the processing of reviews of candidate services by previous users to get dataset as set of keywords as key and get average rating.
2. Then in second phase, it computes similarity between current and previous user by mapping the tuples of keywords using sim function.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

3. The key considered is keywords from keyword set and occurrence as value pair, ratings for keyword will be fed to Map() function .
4. Map() function will collect all the data related to respective review of keyword set and map it i.e. compute occurrence for each same and different keyword.
5. Now the processed data will be fed to Reduce () function for shuffling of occurrence to single computed value for each same and different word. This is helpful to get personalized list finally.

- Now the execution of Exact Similarity Computation is fed to map and reduce function :

1. First phase dose the same processing as in ASC is done.
2. In Second phase, it processes all the reviews of previous user with preference weight of keyword set using TF/IDF measurement by map () function which shuffles the same and different users with set of keywords used.
3. Finally it compares the current and previous user preferences by sim function and generates personalized list for user.

IV. PSEUDO CODE

Step 1: Generate Candidate Keyword list of user's preferences based on previous users reviews collected by stemming and stop words removal.
Step 2: Provide current user with set of keywords for their choice and extract those keywords for further processing.
Step 3: Compare current user and previous user keywords using jaccard coefficient of ASC algorithm using eq. (1).
Step 4: Assign weight to set of keywords captured based on number of occurrence in the reviews and compute TF/IDF for those keywords for further processing.
Step 5: Compare weighted keyword list using cosine similarity function of ESC algorithm eq. (2).
Step 7: Once the comparison phase is completed process ASC and ESC generated keyword list through Map reduce using Mapper and Reducer class.
Step 8: Finally generate personalized list by comparing with some threshold value ∂ .
So as to enhance efficiency system can distinguish positive and negative preferences of user's.

V. SIMULATION RESULTS

The demonstration includes dataset of reviews of hotels. The User based collaborative filtering method is used to generate personalized list to user using KASR algorithm. Here Table I describes the comparison of existing and proposed system as per the UPCC (User based Pearson Coefficient Correlation), KASR-ASC and KASR-ESC. The UPCC, KASR-ASC and KASR-ESC are used to evaluate the result generated using Pearson Coefficient Correlation Constant, jaccard coefficient and cosine similarity function respectively which checks relevancy of keyword list generated during processing of reviews and recommended list of hotels. And this uses DCG (Discounted Cumulative Gain) values for evaluating ideal ranking of recommendation entities to get Top-k recommendation list (where k is ranking measure = 3,5,7).

Approach	Existing evaluation DCG values			Proposed DCG values		
	Top-3	Top-5	Top-7	Top-3	Top-5	Top-7
UPCC	5.09	5.32	5.36	5.09	5.34	5.4
KASR-ASC	7.02	7.78	7.91	7.02	7.88	7.99
KASR-ESC	8.53	9.06	9.11	8.53	9.08	9.15

TABLE I: Comparison of UPCC, KASR-ASC and KASR-ESC approaches with existing and proposed DCG values

In Fig. 1 shows the graphical representation of comparison of recommendation approaches for UPCC, KASR-ASC and KASR-ESC with respect to existing system respectively. Here the comparative value difference is based on DCG values and processing is without distinguishing preferences of users.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

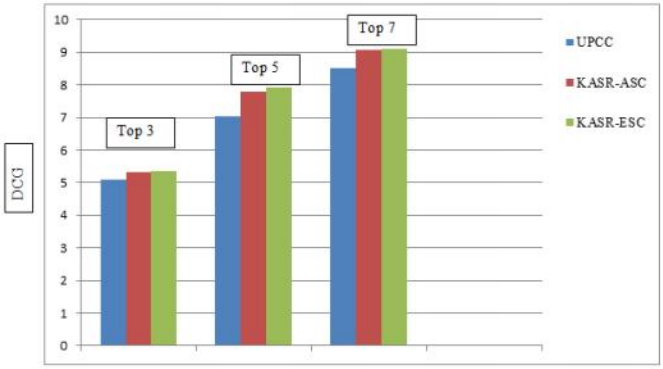

Fig. 1: Comparison between UPCC, KASR-ASC and KASR-ESC in Existing system

In Fig. 2 shows the graphical representation of comparison of recommendation approaches for UPCC, KASR-ASC and KASR-ESC with respect to proposed system respectively which does further processing of distinguishing the positive and negative references of user for enhancing performance.

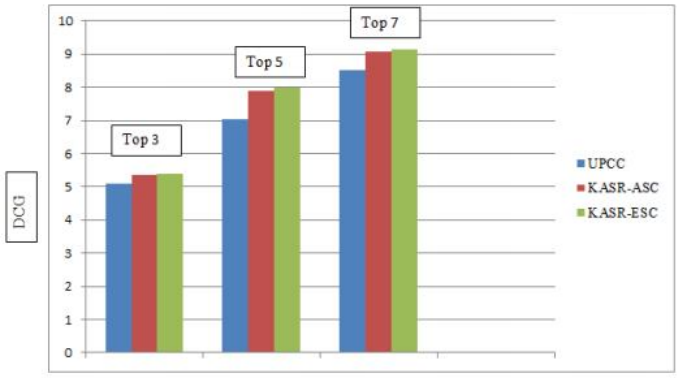

Fig. 2: Comparison between UPCC, KASR-ASC and KASR-ESC in Proposed system

VI. CONCLUSION AND FUTURE SCOPE

The KASR method and user based collaborative filtering algorithm is used for generating personalized user choice. In this paper, we have proposed the method which focuses on distinguishing positive and negative preferences of users for enhancing performance of system. The different datasets are also tested for KASR method for various applications like Hotel reservation system. The Big data in handled in this system using Map Reduce framework for efficient result.
In future scope, it will be useful to work with cross-domain automatic semantic classification for KASR method. Since semi supervised dataset is used in cross-domain environment will be used to generate dynamic or automatic keyword candidate set. Then the system can be tried for processing with Map Reduce.

VI. ACKNOWLEDGE

The authors would like to thank the researchers as well as publishers for making their resources available and teachers for their guidance. We are thankful to the authorities of International Journal of Innovative Research in Computer and communication on Engineering and concern members for their guidelines and support.

REFERENCES

[1] Shunmei Meng, Wanchun Dou, Xuyun Zhang, Jinjun Chen, Senior Member, IEEE, "KASR: A Keyword-Aware Service Recommendation Method on MapReduce for Big Data Applications", IEEE Transaction on Parallel and Distributed Systems, TPDS-2013-12-1141.
[2] Lisette García-Moya, Henry Anaya-Sánchez, and Rafael Berlanga-Llavori Universitat Jaume I , "Retrieving Product Features and Opinions from Customer Reviews" , Published by the IEEE Computer Society ,2013.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

[3] Zheng-Jun Zha ,Member, IEEE, Jianxing Yu, Jinhui Tang,Member, IEEE, Meng Wang, Member, IEEE, and Tat-Seng Chua , "Product Aspect Ranking and Its Applications" , IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 5, MAY 2014.
[4] Danushka Bollegala ,Member, IEEE, David Weir, and John Carroll , "Cross-Domain Sentiment Classification Using a Sentiment Sensitive Thesaurus" , IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 8, AUGUST 2013.
[5] Anindya Ghose and Panagiotis G. Ipeirotis, Member, IEEE, "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics", IEEE Transactions On Knowledge and Data Engineering, Vol. 23, no. 10, October 2011.
[6] Brandon A. Beemer and Dawn G. Gregg, "Dynamic Interaction in Decision Support: Effects on Perceived Diagnosticity and Confidence in Unstructured Domains", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, VOL .43, NO.1, JANUARY 2013.
[7] Kleanthi Lakiotaki and Nikolaos F. Matsatsinis, Technical University of Crete Alexis Tsoukiàs, Université Paris Dauphine, "Multicriteria User Modeling in Recommender Systems" , Published by the IEEE Computer Society, 2011.
[8] Giuseppe Di Fabbrizio, Ahmet Aker, and Robert Gaizauskas, University of Sheffield, "Summarizing Online Reviews Using Aspect Rating Distributions and Language Modeling" , Published by the IEEE Computer Society, 2013.
[9] Xiaohui Yu, Member, IEEE, Yang Liu, Member, IEEE, Jimmy Xiangji Huang, Member, IEEE, and Aijun An, Member, IEEE, "Mining Online Reviews for Predicting Sales Performance: A Case Study in the Movie Domain", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 4, APRIL 2012.
[10] Pravin Jambhulkar1, Smita Nirkhi 2, "A Survey Paper on Cross-Domain Sentiment Analysis", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 1, January 2014.
[11] N.Manjunathan, "Cross-Domain Opinion Mining Using a Thesaurus in Social Media Content", International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 2, Issue 5, May 2014.
[12] http://en.wikipedia.org/wiki/Apache_Hadoop.
[13] http://en.wikipedia.org/wiki/Big_Data.

BIOGRAPHY

Hemlata Kardas is pursuing Masters of Engineering in Computer Science at Rajarshi Shahu College of Engineering and Research, JSPM NTC Pune-India. She obtained Bachelor of Engineering in Computer Science in 2011 at BMIT, Solapur-India. Her area of interest is Web Mining, Information retrieval, Cloud Computing and Hadoop technology. She has presented the conference paper for IEEE Explore on Cloud Computing and Hadoop in 2014 and also published International Journal Papers based on Web mining and Information Retrieval.

Sulochana Sonkamble is HOD of Computer Science Department at Rajarshi Shahu College of Engineering and Research, JSPM NTC Pune-India. She obtained Bachelor of Engineering, Masters of Engg. And PhD in Computer Science Shri Guru Gobind Singhji Institute of Engineering and Technology, Swami Ramand Tirth Marathwada University, Vishnupuri, Nanded-India.