ABSTRACT
Nowadays, data about products, suppliers and consumer experiences on products have been exploding in consumer world in the form of reviews. In order to make use of data, Service recommendation method can be used to extract useful information to the consumer and also to categorize products based on customer preferences. Services are usually recommended based on past user experiences and products are ranked based on ratings from users. In the proposed system, services are recommended based on diverse user preferences by examining the positive and negative aspects in review using valence and arousal algorithm. Initially the review in the form of CSV file is tagged, noun and verb are extracted. Hadoop, a distributed and parallel computing platform is employed for processing the CSV file.

Keywords—Service recommendation, Hadoop, MapReduce, Valence and arousal.

I. INTRODUCTION
In consumer world, the amount of data about product, consumers experience on product, suppliers has been exploding through millions of reviews given by individual user based on their own experiences on the product. Users on various E-Commerce platforms will continue to increase the exponential growth. The user generated information of the E-Commerce platform includes both structured and unstructured data which fall into a category of Big Data analysis. Today’s consumers have evolved beyond being merely ‘buyers’. So, more insights information is necessary from the consumer prospective to buy a product. Service recommendation method usually recommends products to consumer on a common preference. Hence consumer with different preferences is given with same recommendation. Traditional service recommender systems often suffer from scalability and inefficiency problems when processing or analyzing such large, scale data. Current recommendation methods usually can be classified into three main categories: content-based, collaborative, and hybrid recommendation approaches. Content-based approaches recommend services similar to those the user preferred in the past. Collaborative filtering (CF) approaches recommend services to the user that users with similar tastes preferred in the past. Hybrid approaches combine content-based and CF methods in several different ways. In CF based systems, users receive recommendations based on people who have similar tastes and preferences, which can be further classified into item-based CF and user-based CF. In item-based system [4], the predicted rating depends on the ratings of other similar items by the same user. While in user-based system [5], the prediction of the rating of an item for a user depends upon the ratings of the same item rated by similar users. And in this work, we will take advantage of a user-based CF algorithm to deal with our problem.

II. RELATED WORK
A Keyword-Aware Service Recommendation method, named KASR [1], aims at presenting a personalized service recommendation list and recommending the most appropriate services to the users effectively. Specifically, keywords are used to indicate users preferences, and a user-based Collaborative Filtering algorithm is adopted to generate appropriate recommendations. To improve its scalability and efficiency in big data environment, KASR is implemented on Hadoop, a widely-adopted distributed computing platform using the MapReduce parallel processing paradigm. Recommender system is a subclass of information filtering system which attempts to give the guidance to the users regarding the useful services based on their personalized preferences, past behavior or based on their similar likings with other users [3]. It is becoming difficult to capture, store, manage and analyze such big data that affects the service recommender systems with issues like scalability and inefficiency. Also many existing service recommender system provides the same recommendations to different users based on ratings and rankings only, without considering the taste and preference of an individual user.

III. EXTRACTING KEYWORD FROM CSV FILE
Huge Collection of data is retrieved from open source datasets that are publicly available from major Travel Recommendation Applications. The CSV (Comma
separated values) files are read and manipulated using Java API that itself developed by us which is developer friendly, light-weighted and easily modifiable. The CSV Files in distributed Systems are invoked through Web Service Running in the Server Machine of the Host Process through a Web Service Client Process in the Recommendation System. The preferences of active users and previous users are formalized into their corresponding preference keyword sets respectively. An active user can give his/her preferences about candidate services by selecting keywords from a keyword-candidate list, which reflect the quality criteria of the services he/she is concerned about. Besides, the active user should also select the importance degree of the keywords. Initially, HTML tags and stop words in the reviews snippet collection should be removed to avoid affecting the quality of the keyword extraction in the next stage. The Porter Stemmer algorithm is used to remove the commoner morphological and in flexional endings from words in English. The Natural Language Processing comprises tokenizing a sentence or a word, POS (Parts of Speech) tagging, extraction of nouns and verbs, synonym retrieval of extracted keywords using WorldNet dictionary. Each review is chunked and tagged for part of speech. The tagged noun and verb keywords of the review are entered in the keyword-candidate list. In this phase, each review will be transformed into a corresponding keyword set according to the keyword-candidate list and domain thesaurus. If the review contains a word in the domain thesaurus, then the corresponding keyword should be extracted into the preference keyword set of the user.

IV. SIMILARITY COMPUTATION AND USER LIST

The next step is to identify the reviews of previous users who have similar tastes to an active user by finding neighborhoods of the active user based on the similarity of their preferences. Before similarity computation, the reviews unrelated to the active user's preferences will be filtered out by the intersection. If the intersection of the preference keyword sets of the active user and a previous user is an empty set, then the preference keyword set of the previous user will be filtered out. The Natural Language Processing is implemented to analyze the reviews of the previous user. The Big data manipulation from CSV through our own JAVA API enforces developer friendly access.

V. VALENCE AND AROUSAL ASPECT RATING

The chunked reviews of the similar user list are retrieved and the keywords corresponding to the user is analyzed for its Valence and Arousal. Valence means whether the keywords represent the positive or Negative aspect and Arousal signifies the intensity of such aspect. Ratings are given for each domain based on the Valence and Arousal for each user of each hotel. The Overall Hotel Rating is now manipulated by taking average values of each rating of several users of a particular hotel. Now ranking is done for all hotels based on ratings and will be sorted based on Bubble Sort Algorithm to have the most appropriate personalized Recommendation for the User. The Results will be analyzed with graphical views for better understanding.
VI. CONCLUSION

In this paper we have proposed a service recommendation method that addresses both the scalability issue in processing large amount of user generated review and optimizing recommendation for individual user according to their own diversified preferences. To improve scalability and efficiency of proposed method in “Big Data” environments, we have implemented it on a MapReduce framework in Hadoop platform. In addition to rating the service by examining keywords in the review, positive and negative effect that each aspect of the review carries for every keyword is assessed using valence and Arousal algorithm. Thus, the proposed service recommendation method helps in overcoming the scalability problem along with an added advantage of personalized user recommendation for individual user preferences.

REFERENCES


