



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 6, June 2017

A Novel Technique in Review Selection Based on Micro Review

Mumtaj.L.Patel, Prof S.P.Vidhate

PG Student, Dept. of Computer Engineering, VACOE, Sarola Baddi, Ahmednagar, India

Professor, Dept. of Computer Engineering, VACOE, Sarola Baddi, Ahmednagar, India

ABSTRACT: The increase and largely extend of online networking, User can check online review content on various sites such as Yelp.com, foursquare.com etc for checking any restaurant review, product opinions, but Reviews are too long all user cannot read these thousands of review for a item. So that taking item review is very tough task for customers because of writing dummy reviews is also a form of attack, performed to purposefully break an items name. For big review content, users always face the problem of selecting the true reviews. With the recent growth of online web sites for social networking and micro blogging through which users create online communities to share information, thoughts, personal messages and other content has seen an exponential .Micro review provides small review, range limit, cleanliness and faithfulness of dummy reviews. So that by using Micro-reviews user can make the purchase decision on any item.

KEYWORDS: Micro-review, review selection, coverage

I. INTRODUCTION

At the present days there are various web sources to check the online review contents for customers , in that one of the most popular sites that focuses on the reviewing restaurant, Business and share information based on their own experiences such as Yelp.com is a social networking site that has become very popular today. Item reviews are nowadays very much important task for the customer so that they can make decision for purchasing any item or go to any place. The reviewed item includes service providers for example restaurants, hotels etc. But there is also problem of the A bogus review has recently attracted significant interest. Writing bogus reviews is a form of attack, performed to purposefully harm or boost an items name. So with the recent growth of online web sites for social networking and micro blogging through which users create online communities to share their own experiences to other customers about any item. Not shockingly, it has been shown that reviews have incredible effect on the fame of an item. So macro-blocking services that allow users to check-in services. Micro blocking sites also have the ability of boosting the faithfulness of the tips.

II. RELATED WORK

In existing system large corpus of reviews on an item on social networking sites, customer goes through this information filled on web sites because that containing needless words and the length of review is very large. In order to make a purchase decision about an item is very hard task because often hindered by false reviews that fail to capture the true quality of the items attributes. These reviews may be based on insufficient data or may even be fraudulent, submitted to manipulate the items name.

- 1) Large review overloaded on web sites in order to make a purchase decision about an item is very hard task.
- 2) Writing false reviews is a form of attack, performed to purposefully harm or boost an items name.
- 3) Review containing needless words whose content may not be wholly related to the product or item.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 6, June 2017

After defining the problem statement, the survey of journal papers related to defined problem statement has been done as follows. Some of the papers which are related to the defined problem statement are as follows:

A. Selecting a characteristic set of reviews by T. Lappas, M. Crovella, and E. Terzi

In this paper, we formally define the Characteristic-Review Selection problem and prove that it is NP-hard both to solve and approximate. We propose three heuristic algorithms for selecting a characteristic review set, which we evaluate on a wide range of review datasets from different domains. The results indicate that the algorithms are consistently able to find a compact set of reviews that yields a highly accurate approximation of the set of opinions in the corpus. It seeks to preserve the distribution of positive and negative comments. But it cannot be generalized to an arbitrary domain.

Disadvantages:

- 1) Structure data so may not guarantee of appropriate review.
- 2) Not suitable for arbitrary domain.
- 3) Review may be either positive or negative

B. Selecting a comprehensive Set of Reviews by P. Tasaparas, A. Ntoulas, E. Trezi

In this paper, which formulates the review retrieval problem as a maximum coverage problem and tries to select a small number of high-quality reviews having different view-points and covering a maximum number of different aspects reviewed product(+,-). Provide authentic review using TOPQLTY algorithm sorting technique. Problem is based on limited review set.

Disadvantages:

- 1) Uses TOPQLTY algorithm which is based on limited review set.
- 2) Select review from review.

C. Efficient confident search in large review corpora by T. Lappas and D. Gunopoulos,

In this paper, they formalize the Confident Search paradigm for review corpora, then present a complete search framework which, given a set of item attributes, is able to efficiently search through a large corpus and select a compact set of high-quality reviews that accurately captures the overall consensus of the reviewers on the specified attributes, also introduce CREST (Confident Review Search Tool), a user-friendly implementation of our framework and a valuable tool for any person dealing with large review corpora. It is equipped with an efficient method for filtering redundancy. The filtered corpus maintains all the useful information and is considerably smaller, which makes it easier to store and to search. Problem is this system works on artificial review.

Disadvantages:

- 1) System works on artificial review.
- 2) Less efficiency.

D. Selecting a diversified set of reviews by W. Yu, R. Zang, X. He, C. Sha,

In this paper, they proposed an approach to select a small set of representative reviews for each product. Which shall consider both the attribute coverage and opinion diversity. It provides better diversification results especially for selecting smaller sets of reviews.

Disadvantages:-

- 1) Set of reviews should be small.
- 2) Third party opinion product.

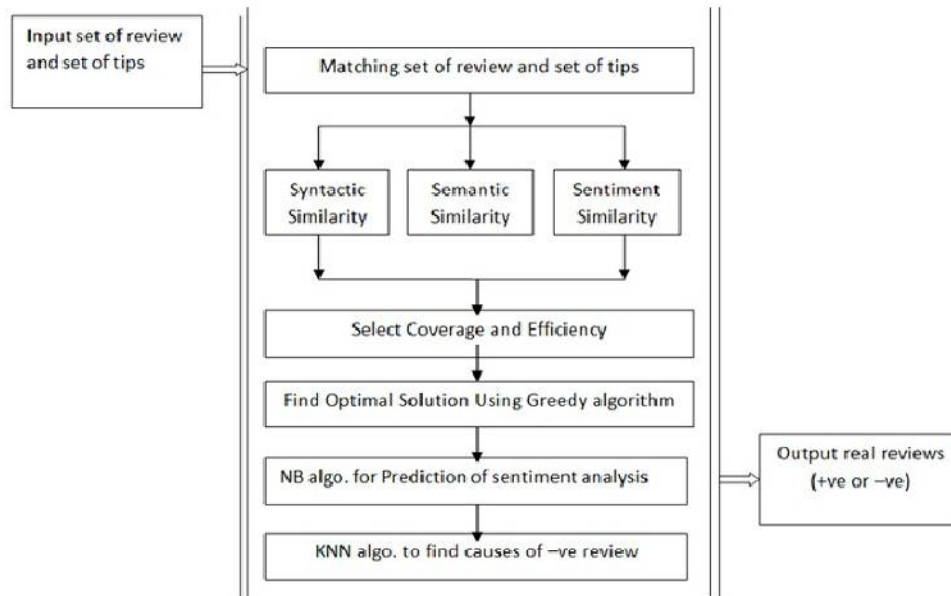
International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

III. PROPOSED SYSTEM



Block Diagram of Proposed System Architecture

Matching Review and Tips

We are given as input a collection of reviews set and set of tips about an item . Our aim is to select the small number of reviews that best covers the set of tips for that we are performing matching function and for matching reviews and stored tips, we consider three types of similarity:-

Syntactic similarity: -

A review sentence and a tip are syntactically similar if they share important keywords.

Semantic similarity: -

A review sentence and a tip are semantically similar, when they are describing the same concept, even if they do not use exactly the same keywords.

Sentiment similarity:-

goal is to predict the sentiment (positive or negative)of a sentence or a tip.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 6, June 2017

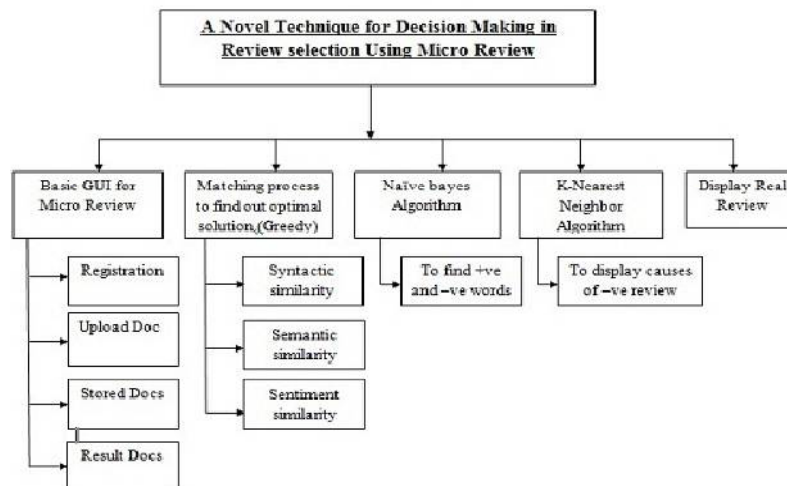


Fig. 2. Breakdown Structure

The breakdown Structure is shown in Figure 2 which mainly focuses on following modules:

Module1: In this phase we create Basic GUI of Big Data Analytics this module consists of basic communication ow between User and System. In this module, User will do registration after registration user can upload documents and also stored results.

Module2: In this module we are giving set of reviews and set of tips as an input and performing matching function on it.we want to select a small set of reviews that best cover the content of the tips. So that we use semantic similarity, syntactic similarity and sentiment similarity .We also uses diferent attributes like coverage and efficiency for selecting real set of review so that we can find out optimal solution by using greedy algorithm.

Module3: In this module we use NB algorithm for prediction of Positive and Negative words and that contains two classes.After matching review and tips we will check the efficiency of both classes.

Module4: In this module we are using k-NN algorithm to show the causes of negative review.

Input Of Module: The input for this will be a set of review and set of tips.

Output Of Module: The result containing Real review either Positive or Negative.

IV. PROPOSED ALGORITHM

A. GREDDY ALGORITHM:-

Step1: Identify Tips.

Step2: Split the sentences of review in separate words.

Step3: Store the words in efficient data structure.

Step4: Store the spitted word of review and tips in efficient Data structure.

Step5: Compare the words of tips with the word of reviews and store the no. of occurrence in DS.

Step6: Find the efficiency and coverage of each words and fit this as input to find semantic analysis.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

B. .NB ALGORITHM:-

Step 1: Convert the data set into a frequency table.

Step 2: Create Likelihood table by finding the probabilities.

Step 3: Now, use Naive Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction.

C. k-NN ALGORITHM:-

In *k-NN classification*, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its *k* nearest neighbors . If *k* = 1, then the object is simply assigned to the class of that single nearest neighbor. In *k-NN regression*, the output is the property value for the object. This value is the average of the values of its *k* nearest neighbors.

V. RESULTS

Screens

Registration for user

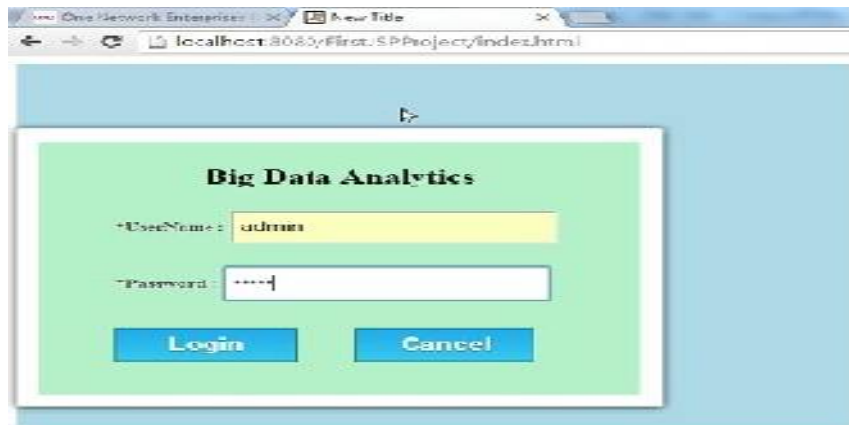


Figure 4.1: Login

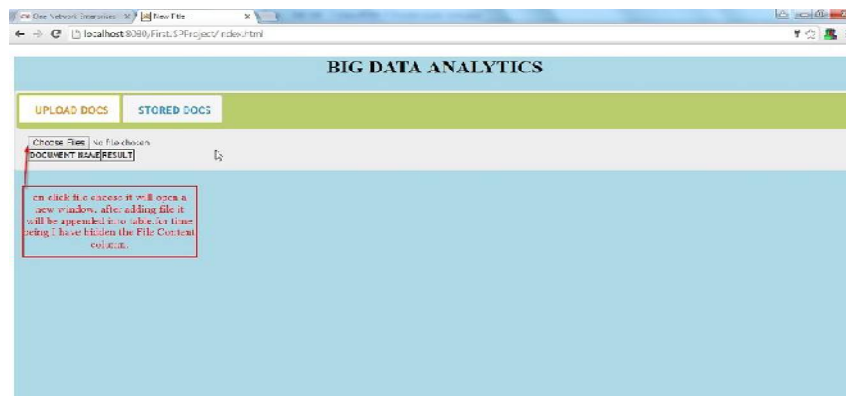


Figure 4.2: Menu

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

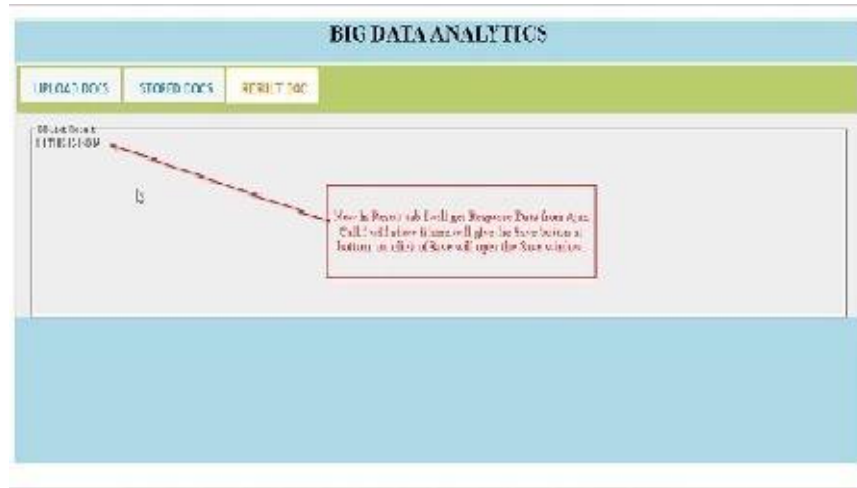


Figure 4.3: Upload Documents

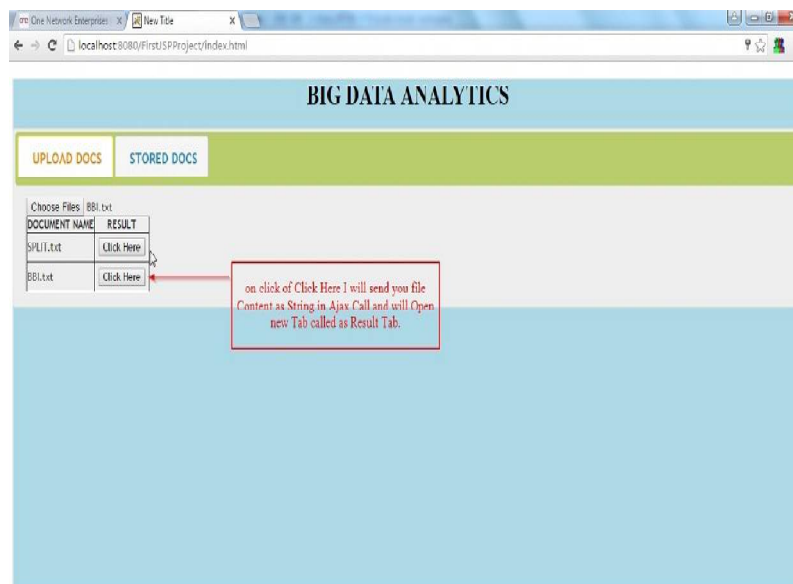


Figure 4.4: Document Uploaded

VI. CONCLUSION AND FUTURE WORK

The proposed system of a novel technique for decision making that help the customers by providing real review on an item.

- By using Micro review we can select small set of review from big.
- System will provide short review either product is good or bad(easy to read for customer)
- Save the customer time.

More work can be done on unstructured data and system will analyze infinite set of data.



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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

REFERENCES

- [1] Thanh-Son Nguyen, Hady W. Lauw, Member, IEEE, and Panayiotis Tsaparas, Member, IEEE Review Selection Using Micro-Reviews.
- [2] T. Lappas, M. Crovella, and E. Terzi, Selecting a characteristic set of reviews, in Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, 2012, pp. 832840.
- [3] T. Lappas and D. Gunopulos, Efficient confident search in large review corpora, in Proc. Eur. Conf. Mach. Learn. Knowl. Discovery Databases: Part II, 2010, pp. 195210.
- [4] P. Tsaparas, A. Ntoulas, and E. Terzi, Selecting a comprehensive set of re-views, in Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, 2011, pp. 168176.
- [5] Selecting a diversified set of reviews by W. Yu, R. Zang, X. He, C. Sha.,
- [6] M. Hu and B. Liu, Mining and summarizing customer reviews, in Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, 2004, pp. 168177.
- [7] Y. Liu, X. Huang, A. An, and X. Yu, Modeling and predicting the helpfulness of online reviews, in Proc. 8th Int. Conf. Data Mining, 2008, pp. 443452.
- [8] X. Meng and H. Wang, Mining user reviews: From specification to summarization, in Proc. ACL-IJCNLP Conf. Short Papers, 2009, pp. 177180.

BIOGRAPHY

Mumtaj L.Patel received the B.E degree in Computer Science Engineering from Bhujbal knowledge city (MET) in 2013 with First class Distinction and Currently perceiving M.E.degree in Computer Engineering from Vishwabharati Academy's College of Engineering, Sarola Baddi, Ahmednagar, (Savitribai Phule Pune University).

Prof. Vidhate .S.P. Currently working as Professor in Computer Engineering from Vishwabharati Academy's College Engineering, Sarola Baddi, Ahmednagar, (Savitribai Phule Pune University).