



Issue Analysis and Summarization using Opinion Mining

Ronak Talati*, Sudeep Shrivastava, Sayali Khade

Computer Science, Vishwakarma, Institute of Information
Technology, Pune, Maharashtra, India

Abstract— *With less than a decade we have seen a drastic change in our industrial working pattern. The market change from service oriented to customer oriented. In the course of change employee affects the most in various way. The work-force make or break the company. To efficiently run the company it is necessary to know various issue faced by employees and find a lucrative solution.*

Opinion mining is extracting users' opinion from user-generated content. Opinion summarization is useful in feedback analysis, business decision making and recommendation systems. In this paper we show how opinion mining is implemented using a web crawler and is useful in issue analysis and summarization at an educational organisation. We try to evolve our system in the following ways: Easy to Implement, Make effective information gathering, Improve decision support system, Support standard business/industry feature.

Keywords— *Opinion Mining, Summarization, Issue Analysis, Sentiment Analysis, Opinion Summarization*

I. INTRODUCTION

In the last decade the evolution of industrial behavior changed swiftly. Over the decade the nature of business techniques improve or rather say mend due to various direct or indirect impact on various working methodology. The impact happened due to various influential reason such as sustainability, globalization. In this scenario we come to conclusion that current working model widen the gap between the requirement and the result product. To evolve as a whole the company need better interaction between their employer and employee relation-ship at various level. This interaction not only clarify doubts but also lead a well satisfied employer-employee relationship. The better the relationship, the better will be the performance on and off the record i.e. an institution can handle the biggest threats more resourcefully and effectively.

Opinion mining is a type of natural language processing for extracting the review of the person about a particular product or a topic. It involves in building a system to collect and examine opinions about the product made in blog posts, comments, reviews or tweets. There are several challenges in Opinion mining. The first is a opinion word that is considered to be positive in one situation may be considered negative in another situation. A second challenge is that people don't always express opinions in a same way. Most traditional text processing relies on the fact that small differences between two pieces of text don't change the meaning very much. In Sentiment analysis, however, "the picture was great" is very different from "the picture was not great". People can be contradictory in their statements. Most reviews will have both positive and negative comments, which is somewhat manageable by analyzing sentences one at a time. However, in the more informal medium like twitter or blogs, the more likely people are to combine different opinions in the same sentence which is easy for a human to understand, but more difficult for a computer to parse[1].

Opinion summarization is a process of generating summary from multiple reviews. It is based on feature selection, feature rating and identifying sentence that contain feature. Opinion mining is a topic in Text mining, Natural Language Processing (NLP), and Web mining discipline. Though user generated contents has proven useful in many applications, challenges still exist in process of opinion mining due to unstructured and noisy data on websites. Everyday creation of user-generated content in a large scale and because it involve user attitude there is need of opinion mining[2]. We therefore use the above technique to analyse issues faced by the user at an educational or a corporate level.

II. LITERATURE SURVEY

Pang and Lee [3] presented survey on sentiment analysis and opinion mining. In that survey they explained opinion oriented information access, challenges, opinion classification and summarization. Mikalai Tsytasarau, Themis Palpanas [4] also have presented Survey on opinion mining. In that survey author explained opinion mining, opinion aggregation and subjectivity analysis. Their study mentioned different work performed on this issue and their comparisons. Earlier sentiment analysis performed on different domain data such as Movie, Products, Restaurants, and Travel etc[2]. Many authors applied opining mining concept to social network data of Facebook, data from messages written by users. Many researches [15] developed sentiment analysis applications on twitter data. Other issues in opinion mining are emotion recognition, opinion spam detection. In [6] author proposed different methods such as machine learning, machine translation and dictionary for sentiment analysis of text in Hindi language. Many researchers used machine learning methods for sentiment analysis that involve training of classifier on datasets and use the trained model for new document classification. Some authors suggested another method such as dictionary of word lexicons [5] [6]. The Dictionary approach is based on a prebuilt dictionary that contains opinion polarity values of words. Many resources are developed

in this domain are datasets , polarity annotated corpus , dictionaries. SentiWordNet is a resource that contains word polarity values which is based on WordNet dictionary. An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

An important part of our information-gathering behavior has always been to find out what other people think. With the growing availability and popularity of opinion-rich resources such as online review sites and personal blogs, new opportunities and challenges arise as people now can, and do, actively use information technologies to seek out and understand the opinions of others. The sudden eruption of activity in the area of opinion mining and sentiment analysis, which deals with the computational treatment of opinion, sentiment, and subjectivity in text, has thus occurred at least in part as a direct response to the surge of interest in new systems that deal directly with opinions as a first-class object.

This survey covers techniques and approaches that promise to directly enable opinion-oriented information-seeking systems. Our focus is on methods that seek to address the new challenges raised by sentiment-aware applications, as compared to those that are already present in more traditional fact-based analysis. We include material on summarization of evaluative text and on broader issues regarding privacy, manipulation, and economic impact that the development of opinion-oriented information-access services gives rise to. To facilitate future work, a discussion of available resources, benchmark datasets, and evaluation campaigns is also provided.

III. OPINION MINING

Opinion Mining also called sentiment analysis is a text extracting technique of finding user's opinion towards a topic. Opinion mining depicts whether the user's view is positive, negative, or neutral about product, topic or an event. Opinion mining involves analyzing user's opinion, attitude, and emotion towards particular topic. This consists of first categories text into subjective and objective information, and then finding polarity in subjective text [4]. Opinion mining can be performed word, sentence or document level. Opinion mining and summarization process involve three main steps, first is Opinion Retrieval, Opinion Classification and Opinion Summarization. Review Text is retrieved from review websites. Opinion text in blog, reviews, comments etc. contains subjective information about topic. Each crawled reviews are treated as a document vector. Reviews classified as positive or negative review. Opinion summary is generated based on features opinion sentences by considering frequent features [4].

I. Opinion Retrieval

Opinion retrieval is a process of collecting reviews text from review websites. Different review websites contain reviews for products, movies, hotels, news etc. Information retrieval techniques such as web crawler can be applied to collect review text data from many sources and store them in database. This step involves retrieval of reviews, micro-blogs, comments etc of user. We should only consider the data which contain subjective data but not the objective data. Reviews are retrieved by query based information retrieval techniques.

II. Opinion Classification

Classification of the review text. The opinions are classified into three categories, positive-neutral-negative. We use a web Crawler to classify data retrieved.

A. Crawler Algorithm

Step 0: Start

Step 1: Get a seed URL

Step 2: Detect all href tag for links and add into a Queue

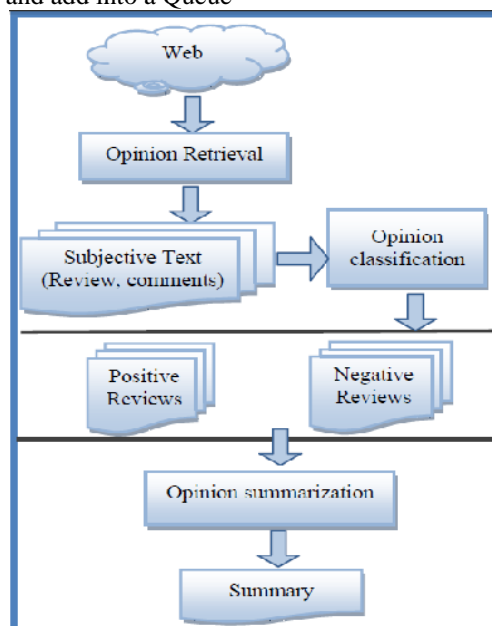


Fig. 1 Opinion Mining

- Step 3: Create a URL for the link
- Step 4: Visit all the URL of the links
- Step 5: repeat step 2
- Step 6: Add all links list into a tree
- Step 7: repeat till all branch urls been crawled
- Step 8: Stop

IV. SUMMARIZATION

Summarization of opinions is a major part in opinion mining process. Summary of reviews provided should be based on features or subtopics that are mentioned in reviews. Therefore, feature extraction and opinion summarization are key issues. The opinion summarization process mainly involve following two approaches.

A. Feature based summarization

This type of summarization involves finding of frequent terms (features) that are appearing in many reviews. The summary is presented by selecting sentences that contain particular feature information. Features present in review text can be identified using Latent Semantic Analysis (LSA) [6] method. For a short summary of product reviews, product features and opinion words associated with it can be presented as summary. Sentences in which feature and opinion words are present are displayed in summary of reviews.

B. Term Frequency based summarization

Term frequency is count of term occurrences in a document. If a term has higher frequency it means that term is more import for summary presentation. In many product reviews certain product features appear frequently and associated with user opinions about it. In this method sentences are scored by term frequency [4]. The summary is presented by selecting sentences that are relevant and which contain highest frequency terms. Opinion Summarization process the review text is preprocessed which involve sentence segmentation and tokenization of sentence in terms. After calculating term frequency of each term, each sentence score and relevance is calculated. As per the compression rate highest scoring and relevant sentences are presented in summary.

C. Text Summarization Based on Fuzzy Logics

A Design of a Fuzzy logic system usually involves selecting membership function and fuzzy rules. The performance of the fuzzy logic system will directly affect by the selection of fuzzy rules and membership functions. The four main components of the Fuzzy Logic were: fuzzifier, inference engine, defuzzifier, and the fuzzy knowledge base. In the fuzzifier section, snappy inputs are translated into linguistic values, using a membership function, which is to be used to the input linguistic variables. After fuzzification, to derive the linguistic values, the inference engine refers to the rule base containing fuzzy IF THEN rules. Finally, the defuzzifier converts the output linguistic variables from the inference to the final crisp values using membership function which represents the final sentence score. The output membership function in the defuzzification step is divided into three membership functions: Unimportant, Average, Important, Which is used to convert the inference engine result into a crisp output to obtain a final score for each sentence. Fig. 2 shows text summarization based on fuzzy logic system architecture. Here fuzzy centroid method is used to calculate the score for each sentence in a document, which is obtained by using generalized triangular membership function which depends on the three parameters a, b, and c. where the parameter a and c are left and right most feet of a triangle and b is the peak of a triangle. Based up on the sentence features and knowledge base the output is obtained as a value from zero to one for each sentence. Such obtained value shows the degree of importance of the sentences in the final summary.

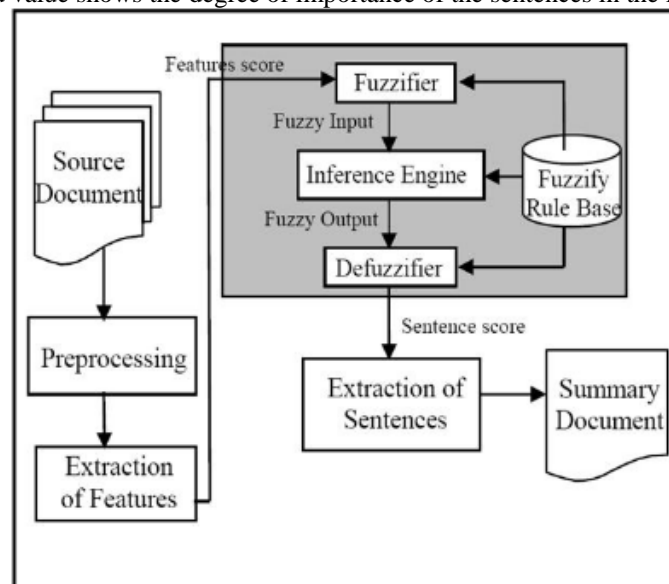


Fig. 2 Shows Summarization Based on Fuzzy Logic

The formula to calculate the fuzzy centroid is given below[8]:

$$f(x, a, b, c) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right)$$

The standard values of Low Medium and High were denoted by a, b, c respectively. The most important part in the Inference Engine is the definition of Fuzzy If Then rules.[8] The most important sentences were taken out from the rules mentioned below, which is based on our features selection:

IF (Indicatorcuephrase is VH) and (Legalvocabulary is H) and (Paragstructure is VH) and (Citation is H) and (Termwght is VH) and (Nameentityrecog is H) and (Absolutelocation is H) and (SentenceSimilarity is VH) and (NoProperNoun is M) and (SentencePosition is H) THEN (Sentence is important)

V. SYSTEM ARCHITECTURE OF ISSUE ANALYSIS AND SUMMARIZATION

I. Web Parsing

In this module, all the URL's linking from the home page(Seed URL) analysis is done and then they are store in a vector. All the sub URLs are stored inside the vector. The page content of all the URLs is parsed.

The opinions given to all issues generated is parsed by the crawler.

II. Preprocessing

The content parsed from the web page is preprocessed again and again. The content is then classified as Top Words, Title Sentences, Numerical Data. Each is given a rank. The content is now classified and opinion mining is completed. The retrieved and tokenized text is passed to the fuzzy logic for summarization.

III. Fuzzy Logic

The preprocessing output works as an input for the Fuzzy Logic part. This summarizes the opinions and gives a final opinion. A summarized opinion to a given issue is generated from all the opinions and reviews given.

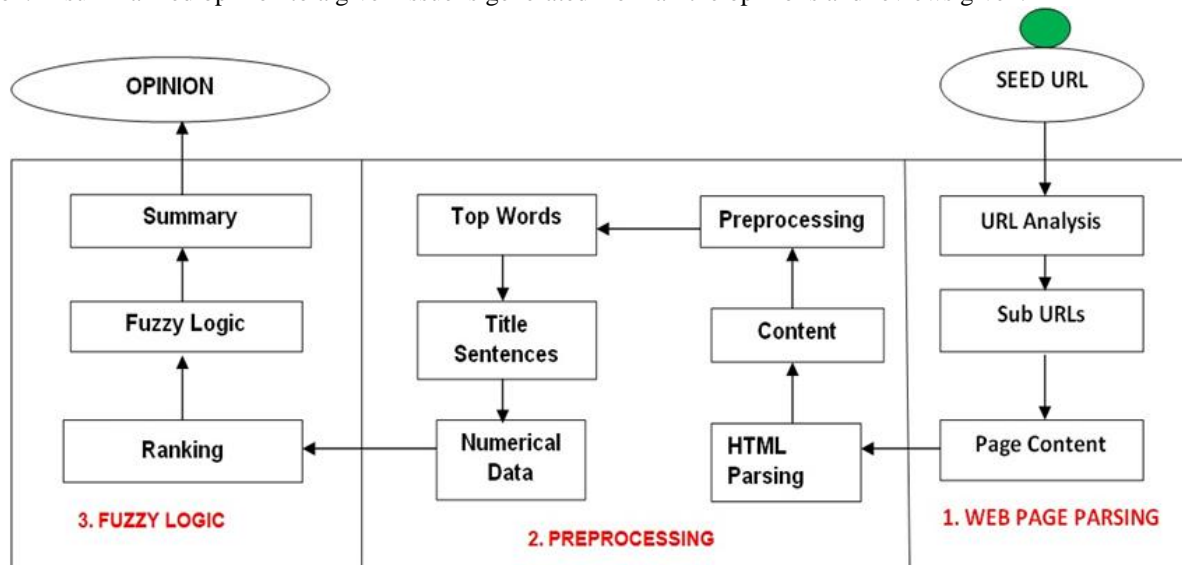


Fig. 3 System Architecture

VI. CONCLUSION

Opinion Mining and text summarization has many real time applications. In the above paper we show how issue analysis and summarization can be implemented using opinion mining using a web crawler and summarize the opinions using the fuzzy logic technique. It is a very efficient and time saving way to get quick opinions from the users and summarize their reviews easily.

In future, better algorithms can be used to increase the performance. Although the techniques and algorithms used for sentiment analysis are advancing fast, however, a lot of problems in this field of study remain unsolved. The main challenging aspects exist in use of other languages, dealing with negation expressions; produce a summary of opinions based on product features/attributes, complexity of sentence/ document , handling of implicit product features , etc. More future research could be dedicated to these challenges.

ACKNOWLEDGMENT

We would like to take this opportunity to express our gratitude towards all the people who have in various ways, helped in our project. We must convey our gratitude to Prof.Mr.S.G.Walunj for being the constant source of inspiration and helping in the project work, and providing encouragement throughout the project. We also thank all our faculty members for steering us through all the phases of the project in a result oriented manner with concern attention.

REFERENCES

- [1] G.Vinodhini RM.Chandrasekaran in Sentiment Analysis and Opinion Mining: A Survey:Volume 2, Issue 6, June 2012 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering
- [2] Vijay B. Raut et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 1026-1030
- [3] Bo Pang, Lillian Lee, "Opinion Mining and Sentiment Analysis",Foundations and Trends in Information Retrieval Vol. 2, Nos. 1–2 (2008).
- [4] Mikalai Tsytarau, Themis Palpanas "Survey on mining subjective data on the web", Data Mining Knowledge Discovery, Springer 2012, pp.478-514.
- [5] Elena Lloret, Alexandra Balahur, José M. Gómez, Andrés Montoyo, Manuel Palomar, "Towards a unified framework for opinion retrieval, mining and summarization" Journal of Intelligent Information Systems Springer 2012, pp.711-747.
- [6] Alexandra Balahur, Mijail Kabadjov, Josef Steinberger, RalfSteinberger, Andrés Montoyo, "Challenges and solutions in the opinion summarization", Journal of Intelligent Information Systems Springer 2012, pp.375-398.
- [7] Alexandra Trilla, Francesc Alias "Sentence-Based Sentiment Analysis for Expressive Text-to-Speech", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 21, No. 2, February 2013, pp.223-233.
- [8] S.SanthanaMegalaeta,/(IJCSIT) International Journal of Computer Science and Information Technologies, Vol.5(1),2014,863-867.
- [9] M. Esther Hannah, T.V.Geetha,Saswati Mukherjee,“Automatic Extractive Text Summarization Based on Fuzzy Logic: A Sentence Oriented Approach, Springer-Lecture Notes in Computer Sceince, P:530-538, 2011.
- [10] S. Rucha, S.S.Apte, "Improvement of Text Summarization using Fuzzy Logic Based method", IOSR Journal of Computer Engineering, P:5-10, 2012.
- [11] Aditya Joshi, Balamurali A R, Pushpak Bhattacharyya, Rajat Mohanty, "C-Feel-It: A Sentiment Analyzer for Micro-blogs", Proceedings of the ACL-HLT 2011, pp.127-132.