



## International Journal of Engineering, Business and Enterprise Applications (IJEBA)

[www.iasir.net](http://www.iasir.net)

### An Empirical Study of Extracting information for Business Intelligence

V.Jayaraj<sup>1</sup> V.Mahalakshmi<sup>2\*</sup>

<sup>1</sup>Associate Professor <sup>2</sup>Research Scholar

<sup>1,2</sup>School of Computer Science & Engineering, Bharathidasan University, Tiruchirappalli-24, Tamilnadu, India

**Abstract:** Sentimental/opinion analysis is an emerging area of research in text mining. Sentimental analysis or opinion mining refers to identify and extract subjective information in source materials. As a response to the growing availability of informal opinionated texts like blog posts and product reviews, comments, forums which is collectively called as user generated contents. A field of sentimental analysis has sprung up in the past decades to address the question what do people feel about certain topic? Bringing together researchers in computer science, data mining, sentimental analysis expand the traditional fact-based text analysis to enable opinion- oriented information systems. This paper provides an overall survey about sentiment analysis or opinion mining related to Business intelligence.

**Keywords:** Opinion mining, Opinion analysis, Text Mining, Business Intelligence.

#### I. INTRODUCTION

Dealing with the ever growing information in the internet opinion mining plays an essential part in our information gathering before taking a decision. Opinion mining is the area of research refers to identify and extract subjective information in source materials. Opinion mining is also referred as sentimental analysis. Opinion Mining concentrates on classifying documents according to their source materials [1]. The main goal of an Sentimental analysis is to determine the polarity of comments (positive, negative or neutral) by extracting features and components of the object that have been commented on in each document .A basic task in sentiment analysis is classifying the polarity of a given text at the document, sentence, or feature/aspect level whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral[2]. As a response to the growing availability of informal opinionated texts like blog posts and product review websites, a field of sentimental analysis has sprung up in the past decades to address the question What do people feel about certain topic? Sentiment classification classifies whether an opinionated document as positive or negative [3]. A text document is classified using a machine learning techniques (Naive Bayes, Maximum Entropy, support vector machines)[4]. A piece of text can be used as an feature or object in opinion mining. The opinion expressed in every document is either direct opinion or comparative opinion. Direct opinion express a target, a person etc. (e.g) I bought an Nokia x2 mobile. Comparative opinion express e.g. laptop x is cheaper than laptop y. Opinion mining task is carried out in the sentence and document levels. Subjectivity/ Sentence level opinion mining is performed by two tasks. Subjectivity classification identifies whether a sentence is subjective or objective. The research in the field started with sentiment and subjectivity classification, which treated the problem as text classification problem. Sentiment classification classifies whether an opinionated document as positive or negative. Subjectivity classification identifies whether a sentence is subjective or objective. Many applications required more detailed analysis because the user wants to know the opinion of others. Let us consider the following example, (1)I bought an galaxy mobile 4 days ago. (2) It was an beautiful phone.(3)The touch screen was really superb. (4)The Voice excellence was also good. (5)However, my father was fight with me as I didn't inform him before I bought it.(6)He felt that the mobile was too costly, and wanted me to return it to the shop.

The question is: what we want to know from this review? There are several opinions in this review (2),(3),(4) express positive opinion, while (5) and (6) express negative opinion. The opinion in sentence (2) is on galaxy mobile, (3) is on touch screen and (4) voice excellence are the features of galaxy mobile. Sentence (6) is on the cost of a galaxy mobile .This is an important place to understand the users are interested on other opinions, but not on all. With this example in mind, we can define opinion mining ,an opinion can be expressed as target, opinion holder, opinion and orientation, direct opinion, comparative opinions.Finding the relevant information about companies from the multiple sources on the web has become increasingly important for business analysts. To get an accurate result of a business entity, text mining tools have been used. With the appropriate tools, company analyst would have to read thousands of reports, news articles etc.This paper is organized as follows: In section 2, various research works has to be analyzed in order to enhance our work. In section 3, our discussion has been described in details. Finally, the paper is concluded by summarizing the work

## II. RELATED WORK

Wenhg Zhang et al., [7] identified the weakness of the product by using weakness finder algorithm. The algorithm extract the implicit and Explicit features using morpheme based method and hownet based method to determine the polarity of each sentence. The weakness of the product has to be identified because to know the unsatisfication of the customers and compared with the competitors product reviews to improve their product weakness. Guang Qiu et al.,[8] proposed an advertising strategy DASA to promote advertisement and then to identify the negative review of the customers. These approaches uses pre-set rules, and also design an prototype system for the users.

Shumin Zhou, et al.,[9] proposed an architecture to connect the government and the people .The customers may post their opinion by mobile or internet named as information collection channels. The architecture is named as people opinion collection processing. the dataflow process starts and then collect and processing pocp This POCP promotes to build the harmony society.

To evaluate the extraction system, we use traditional metrics for information extraction Chinchor et al., [10] calculate the precision, recall, and F-measure values . Precision measures the number of correctly identified items as a percentage of the number of items identified. It measures how many of the items that the system identified were actually correct, regardless of whether it also failed to retrieve correct items. The higher the precision, the better the system is at ensuring that what is identified is correct. Recall measures the number of correctly identified items as a percentage of the total number of correct items measuring how many of the items that should have been identified actually were identified. The higher the recall rate, the better the system is at not missing correct items. The F-measure is often used in conjunction with Precision and Recall, as a weighted average of the two usually an application requires a balance between Precision and Recall.

Horacio Saggion et al., [11] finding the relevant information about companies from the multiple sources on the web has become increasingly important for business analysts. To get an accurate results of an business entity, text mining tools have been used. With the appropriate tools, company analyst would have to read thousands of reports, news articles etc. M.Rushdi Saleh et al.,[12] Opinion mining is receiving more attention due to the increase of blogs, forums, websites etc, support vector machine has been used for testing the dataset and using several weighted schemes.

In this work, Support Vector Machines have been applied in order to classify a set of opinions as positives or negatives.svm has achieved good results in opinion mining.svm has also been successfully achieved in many classification tasks. SVM has applied with different features in order to test how the sentiment classification is affected. Different weighting schemes (TFIDF,BO) and n-grams techniques are used. By using the svm tool sentiment orientation classification was fulfilled. Symbolic approaches and machine learning techniques are extended in order to attack the classification of reviews .

Dietmar Gräbner et al., [13] proposed a lexicon based approach to classify the customer reviews based on sentimental analysis .when the precision and recall values exceeds the given baseline of our approach with the algorithm for sentimental analysis proved to be successful. Generate a reliable classification approach of customer reviews by applying lexicon based sentimental analysis. Three steps to be carried out to create an lexicon 1.build an lexicon with semantic orientation 2. Create an sentimental analysis based lexicon to generate classification reviews 3.classification results are evaluated with quantitative ratings.

Zhongwu Zhai et al.,[14] proposed an several methods have been proposed to extract product features from the reviews. very limited work has been done in the clustering. Lexical similarity can be used in clustering but it was not still accurate because with very high similarities are reliable. so to overcome these problems proposed an semi supervised learning. For semi supervised learning, use the EM algorithm formulated in which is based on NaïveBayes classification. EM algorithm performs much better when compare to the other algorithm. Due the poor performance of the unsupervised methods an EM algorithm based on Naïve Bayes classification is adapted to solve this problem. After a semi supervised method applied then connect feature expressions using sharing words, and then merge components using lexical similarity and select the leader components as labeled data. Alexandra Balahur et al.,[15] proposed an method to evaluate an used generate content. In order get knowledge from user generated content, automatic methods must to be developed.To multi document summarization of opinions from blogs, forums etc.

Vast different approaches have been used to identify the positive and negative opinions and then summarize the opinions. The aim of the work is to study the manner in which opinion can be summarized, so that they obtained summary can be used in real-life applications e.g marketing, decision- making. Business Intelligence (BI) is a process for increasing the competitive advantages of a business by intelligent use of available information collection for users to make wise decision [16], [17]. It was well known that some techniques and resources such as data warehouses, multidimensional models, and ad hoc reports are related to Business Intelligence [18]. Although these techniques and resources have served us well, they do not totally cover the full scope of business intelligence [19].

### III. DISCUSSION

Sentimental analysis play a vital role in business intelligence and also organizations. Decision making is big issue always in many organizations. 80% of information in companies are unstructured data. To get the relevant information from that unstructured information plays a main role for the analyst. Information Retrieval concepts play a main role in classifying unstructured data. By using this technique our work can be extended and the meaningful data can be retrieved. People get the others opinion to make some decision about product or services by this ways.

- Finding opinions while purchasing the product
- Finding the opinion of the competitor products
- Finding opinions on tender result

Finally getting an relevant information about product or services plays an main role in an organizations. The core objective of the paper is to develop a methodology to mine the useful information from the unstructured textual content in order to improve the business intelligence. The mining process can be achieved by new emerging technology, which is variant from data mining. With the help of text mining, the user can able to discover previously unknown knowledge in text, by automatically extracting information from different written resources developed in natural languages. It can be now familiar because of its approaches to information management, research and analysis. Thus, text mining is the extension of data mining and obtains the goal of extracting meaningful data from different sources of textual documents. In data mining, the collection of data is stored in the repository known as Data Warehouse. Likely, in text mining, the collection of documents is stored in the repository known as Document Warehouse. From this Document Warehouse, the text has to be extracted using text mining.

### IV. CONCLUSION AND FUTURE WORK

In this literature survey paper it is observed that opinion mining play a vital role to make decision about product or services. Finding the relevant opinions expressed on the web, classifying them and filtering only the positive opinions is not helpful enough for the users. They will still have to sift through thousands of text snippets, containing relevant, but also much redundant information. Many organizations are carry out more research in unstructured data. To get the relevant information text mining and information retrieval concepts has been utilized. The work can be further extended to areas like neural networks, XML data information retrieval. In XML retrieval by using configuration techniques a data retrieval time can be optimized.

### V. REFERENCE

- [1] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques", In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 79–86, 2002.
- [2] Dietmar Gräbner, Markus Zanker, Günther Fliebl and Matthias Fuchs "Classification of Customer Reviews based on Sentiment Analysis" In 19th Conference on Information and Communication Technologies in Tourism (ENTER), Springer, Helsingborg, Sweden, 2012.
- [3] Turney, P, "Thumbs Up or Thumbs Down? Semantic orientation Applied to Unsupervised Classification of Reviews", ACL '02, 2002.
- [4] Mital K. Dalal, Mukesh A. Zaveri "Automatic Text Classification: A Technical Review" In International Journal of Computer Applications (0975 – 8887) Volume 28– No.2, August 2011
- [5] Kateryna Rybina "Sentiment analysis of contexts around query terms in documents" in technical universitat Dresden, October 2012
- [6] Pang Bo, and Lee Lillian. *Opinion Mining and Sentiment Analysis*. 2008.
- [7] Wenhao Zhang, Hua Xu, Wei Wan "Weakness finder : Find Product Weakness from Chinese reviews by using aspect based sentimental analysis" in Expert systems with application 2012
- [8] Guang Qiu, Xiaofei He, Feng Zhang, Yuan Shi, JiaJun Bu, Chun chen "DASA: Dissatisfaction –oriented Advertising Based on sentimental Analysis" in Expert Systems with Applications 2010
- [9] Shumin Zhou, Jumei Ai, Congnian Xu, Bin Tang "The collection and processing platform of the peoples opinion Based on SMS and Internet" in IEEE 2007
- [10] Chinchor, N. (1992). MUC-4 Evaluation Metrics. In Proceedings of the Fourth Message Understanding Conference, pp. 22–29.
- [11] Horacio Saggion "Extracting Opinions and Facts for Business Intelligence" <http://www.nist.gov/tac/>
- [12] M. Rushdi Saleh, M.T. Martín-Valdivia "Experiments with SVM to classify opinions in different domains" in Expert Systems with Applications 38 (2011) 14799–14804
- [13] Dietmar Gräbner "Classification of Customer Reviews based on Sentiment Analysis" in 19th Conference on Information and Communication Technologies in Tourism (ENTER), Springer 2012
- [14] Zhongwu Zhai "Clustering Product Features for Opinion Mining" University of Illinois at Chicago.
- [15] Alexandra Balahur "Challenges and solutions in the opinion summarization of user-generated content" © Springer Science+Business Media, LLC 2012
- [16] B. de Ville, "Microsoft Data Mining: Integrated Business Intelligence for e-Commerce and Knowledge Management", Boston: Digital Press, 2001.
- [17] P. Bergeron, C. A. Hiller, "Competitive intelligence", in B. Cronin, Annual Review of Information Science and Technology, Medford, N.J.: Information Today, vol. 36, chapter 8, 2002.
- [18] Bhujade, Vaishali, "Knowledge Discovery in Text Mining Technique using Association Rules Extraction", Computational Intelligence and Communication Networks (CICN), International Conference on oct. 2011.
- [19] M. J. A. Berry, G. Linoff, "Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management", Wiley Computer Publishing, 2nd edition, 2004.