Special Issue Published in International Journal of Trend in Research and Development (IJTRD), ISSN: 2394-9333, www.ijtrd.com

Opinion Mining on E-Commerce: Need of the Hour

J Soonu Aravindan
Department of Computer Science, Bharathiar University, Coimbatore, TamilNadu, India

***Abstract*:** Web mining is becoming hot topic in the area of E-business research and this web mining research relates to different research communities such as Machine Learning, Information retrieval and Artificial Intelligence. It always been a challenging task to understand the trends and extract the opinions about the product in the wide diverse social media data. Thereby this creates the need of automated and real time user emotion reading to generate opinions. This paper is an effort to analyze the recent diversity in web mining with the user emotion analysis. We explored the role of emotion mining for E-business. The survey has been made to categories the recent approaches of emotion analysis, although the emotion analysis is considered as much complicated problem in literature,

Keywords: *Emotion analysis; Opinion Mining; Behavioral Tracking; E-Business; Web mining*

I. INTRODUCTION

Web mining is the application of data mining technique which is used for finding interesting patterns and discovering knowledge from World Wide Web (WWW). Web mining can be broadly divided into three categories. Web Usage mining, Web Content Mining and Web Structure Mining. Due to rapid growth of web content, each and every second we are flourished with data but we are lagging in knowledge. Extracting useful and interesting information from internet becomes a challenging job for the user. In order to consume the browsing time of a user, search engines where introduced, which results in faster extraction of information.

In recent years, number of online users has increased widely with their willingness to participate in social interactions. E-commerce, Social Networking websites allows the user to share their emotions in the form of textual comments instantly. Opinions can be generated based on user emotions, which helps to understand user behavior towards the content. It predominantly supports for document categorization. Several studies have been carried out to automatically predict the most feasible emotions for documents.

Opinion is a judgment held with confidence but not authenticated by proof. It is very essential when it's time to make a decision. When the decision involves spending time and money to buy products, users often trust on past reviews on the products by other users.

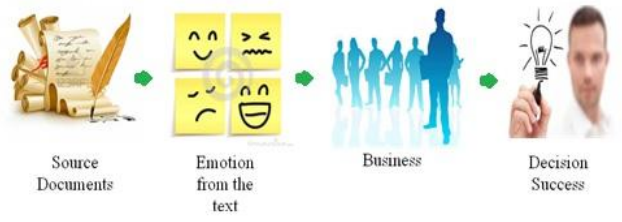

Figure 1: Emotion mining for E-Business

Nowadays, Business Analysts have access to public media where opinions about companies, products, and strategies are expressed in unstructured form. Mining information from public sources is of great importance to many business intelligence applications such as credit rating or company reputation. A text not only conveys informative contents but also shows the attitude and emotional state of a user. In the recent years, research related to user emotions and opinions are emerging rapidly.

The rest of the paper has been organized as follow: Section 2 explains the E-business and relation with the emotion or sentiment analysis for a success. The emotions of the users and various studies related to the emotions have been discussed in section 3. Section 4 brings out the various approaches of opinion mining of recent years. Section 5 gives the advantages and usage of emotion analysis in e-business. Finally the paper is come to a conclusion in section 6.

II. E-BUSINESS

Online shopping is the selling of product and services via internet which is accessible at anytime from anywhere in the world. E-Marketing focuses in adapting the current marketing strategies into the web environment. It involves the formation of the E-business, the project itself, the web contents, analytical measures and advertisement. Online promotion is one of the forms of usage of internet in e-marketing where it brings the advertisement message to the targeting groups, quickly and cost effectively. Customer's personal requirement can be analyzed in terms of service, product or information. To meet the user's personal requirement it requires knowing and understanding the needs, expectation and behaviors of the user groups. Apart from the traditional marketing, e-marketing should exploit the data given by the users while they are navigating through the web site. Thus, the user data always remained as a real treasure for e-business. To exploit the user data to maximum, it requires a good e-business models, data warehousing and especial data mining techniques.

The internet is fast medium where the competition in business is also high. The customer groups in such environment are not static. The interest of customer group will change quickly and customer can move away from one provider to another. Good web presence will develop a strong relationship with the customer which will attract the customer to visit the web site frequently for their product and service needs. The marketing strategy for the e-business implies some of the traditional principles with additional strategies. The online business offers transaction data, server data, and web Meta data that can be used to solve and automate the tasks successfully. The trace data leaved by the users can be used to understand the customer needs, desire and demands. In return to which the Web presence of the business can be improved.

III. EMOTION

Emotion is a complex observable fact for which there is no generally accepted definition was ever given. K. Scherer et al. in his work has given a commonly used definition for the emotion as an episode interrelated, synchronized change in the state of the five organismic subsystems in response to assessment of the external or internal stimulus event as relevant to major concerns of the organism [1] [2]. The organismic subsystems with their major substrata are the Information processing (CNS- Central Nervous System), support (CNS, NES- Neuro-Endocrine System, & ANS- Automatic Nervous System), executive (CNS), action (SNS Somatic Nervous System) and Monitor (CNS)[3]. The affect related phenomena's like feeling, moods, sentiment, energy and tension etc., were

been studied in detail by the disciplines like philosophy or psychology. In recent years, Science and technologies has shown an interest in building business strategies to develop a software system that automates the processing of systems with the knowledge of these affects. To accrue the benefits from the knowledge of this phenomenon many inter- disciplinary methods have been proposed using the existing theoretical models as basis for engineering computation.

A. Source of Emotion

1) Sentiments in words.

A word can be a noun, verb, adjective, etc. in order to classify them into sentimental words certain approaches has to be followed. One of the basic approaches is to collect the words that are sentimental to human knowledge and manually classify them into positive or negative polarity. Meanings of

positive word are mostly positive. For example: synonyms for 'Excellent' is 'Outstanding', 'Brilliant', 'Superb', 'Exceptional' and meanings of negative word are mostly negative [4]. For example: antonyms for 'Fraud' is 'Scam', 'Con', 'Fake'.

2) Sentiments in subjective expressions.

Expressions can be either implicit or explicit. Sentiments always involve user emotions implicitly or explicitly. Example for implicit expression is: "We should decrease our dependence on petrol". Example for explicit expression is: "I like chocolates". Polarity can be applied for both the expressions [4].

3) Sentiments in subjective sentences.

A sentence is a collection of words grouped together to form a meaning. In an e-business environment, during product or movie review, opinions of the users are mostly typed as sentences and not just as single word. Classifying the sentence as subjective or objective is important. Objective sentences have to be excluded during analysis. Consider the example: "villan in this film tries to protect his good name" even though this sentence contains the word "good". It tells nothing about user's opinion about the film. And so this kind of irrelevant text has to be excluded [5].

4) Sentiments in Topics.

Topics are nothing but subject what is being talked about. Consider the example: A manufacturing company introduces a brand new mobile in the market with multiple features like hi-speed internet, high resolution camera, stylish looks, etc. at a low cost and the overall review for this mobile is positive. Suppose if the mobile has a complaint regarding the batter backup, it affects the sale. Even though customer likes the mobile he might be not satisfied with the battery backup. To the manufacturers, these individual flaws and strengths are equally valuable than the overall satisfaction of customers. So extracting sentiments about a specific topic plays a vital role in sentiment analysis [6].

B. Classification of Emotion detection in text

The method used to spot the emotional phenomena in the text is classified as follows.

1) Human Reaction Based Models/Methods.

Previous approaches in 80's and 90's to spot the affect in text including the human reactions according to their needs and desires. The few of the most commonly known among them are discussed below in the section.

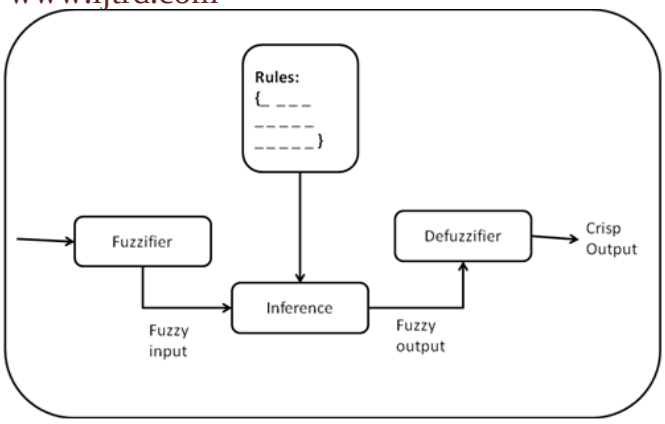

Figure 2: The Fuzzy Inference System

FIS based on fuzzy rules has been applied to numerous engineering applications such as control, signal processing, and pattern classification problems [7] [8]. The Fuzzy logic based methods as shown in figure 2, where the principals of fuzzy are used to recognize the emotion in the data. The classifications is done for each basic emotion Ei, (i = 1, ..., N,) a separate rule set is generated by an adapted fuzzy grid method. Each rule takes the fuzzified features Fj ,(j = 1, ...,K,) as input and produces a fuzzy emotion value as output. There are different works done where the fuzzy is used for the reason of its improved classification and decision systems by allowing the use of overlapping class definition and improves the interpretability of the results by providing more insight into the classifier structure and decision making systems [9].

Emotion is detected in text using the lexical similarity. Here the lexicon used in text is checked for the described emotions. The emotions of the text are classified based on the range of the relation between the classified emotion set and the lexicon. SentiWordNet is one of the examples of such kind of emotion detection method [10]. In another method the emotion is detected with affective Key-WordNet phenomena [11]. The method is used to check the appropriateness of using existing lexical resources or keywords.

The machine learning method using the term frequency is another type of the human reaction based model. Machine learning can be organized as supervised learning, unsupervised learning, semi-supervised learning, transduction, reinforcement learning, learning to learn and developmental learning [12] [13].

Among the above all the machine learning and the affective keyword detection mechanisms are the most widely used approach in emotion detection for their ability of easy adaptability cross domains and languages.

2) Pattern and rules Based.

The designed framework for fuzzyfication of input variables for fuzzy classifier based on statistical analysis of emotions expressed in video records of standard Cohn-Kanade's and Pantic´s MMI face databases [14]. Ying Chen et. al. in his work has proposed a multi label model which not only detects but also based on the linguistic analysis, creates two set of linguistic patterns. The manually generated and the automatically generalized patterns are designed to extract general source expressions or specific constructions for emotion causes [15].

3) Affective reactions based on Common Sense Knowledge.

Senti-Computing is one of the methods used under the category of emotion detection where the affective reactions are detected with the common base of knowledge of Common sense [16]. Liu et. al. in his work has given several emotion detection mechanisms using the knowledge of common sense.

IV. OPINION

Opinion Mining is also known as Sentiment Analysis which is used to identify and extract subjective information from textual contents. A text not only conveys informative contents but also shows the attitude and emotional state of a user. The sentiments found in the reviews, comments or feedbacks can be classified into two categories: positive or negative, or into an n-pointer scale: very good, good, fair, bad, and very bad. Sentiments can be obtained from the following sources.

A. Tools Used in Opinion Mining

Four tools have been discussed here to track the opinion from the user generated contents.

- Review Seer tool [17] - Works done by the aggregation sites are automated by this tool. The Naive Bayes classifier approach is used to collect positive and negative polarities for allocating a score to the mined feature terms.
- Web Fountain [18] - This tool extracts the product features using the beginning definite Base Noun Phrase heuristic approach.
- Red Opal [19] - Based on the features, this tool enables the users to determine the opinion orientations of products. Based on the features extracted from the customer opinion, it assigns the scores to each product.
- Opinion Observer [17] - This is an opinion mining system for analyzing and comparing opinions on the internet using user generated contents. Opinions of the product are showed as feature by feature in a graph format.

B. Opinion Mining Approaches

1) Natural Language Processing (NLP).

NLP involves the interaction between human language and computers. Nowadays, NLP algorithms are based on Machine Learning. In data mining, Decision trees are used for classification and rules are generated. Doing this manually is time consuming and hence classification algorithms are developed for automatic generation of rules where data's were trained using neural networks and when a new data arrives, system will automatically classify with the help of training set. Automatic learning procedure can produce models for unfamiliar inputs. In sentiment analysis, NLP is mainly used for Text analysis. Some of the tasks of NLP in sentiment analysis are Stemming, POS- Part Of Speech, Coreference Resolution, Negation Handling, Word Sense Disambiguation, Terms and their frequency, Sentiment words and phrases, Rules of opinions, Sentiment shifters, and Syntactic dependency. There are several tools available for extracting sentiments. Some of the NLP tools include LingPipe, OpenNLP, Stanford Parser, POS Tagger, OpenFST, NTLK, Opinion Finder, Tawlk/osae, GATE, textir and NLP Toolsuite.

2) Machine Learning Algorithms.

Machine learning is a branch of Artificial Intelligence deals with construction and study of systems that can learn automatically from the data. It mostly deals with construction and representation. Machine learning can be organized as Supervised learning, Unsupervised learning, Semi-supervised learning, Transduction, Reinforcement learning, Learning to learn and Developmental learning. Mostly used machine learning algorithms for Sentiment analysis are SVM- Support Vector Machines, Naïve Bayes, Decision Trees, Logistic Regression, Maximum Entropy, TiMBL, Rule Generators, Genetic Algorithms, Boosting Algorithms, and MLP.

3) Unsupervised Learnings.

Learning from unlabeled data is called unsupervised learning. Sentiment words and phrases may be used to classify sentiments in unsupervised manner. Peter D. Turney has presented a unsupervised learning algorithm for classifying user reviews where classification is predicted by average semantic orientation of the phrases that contains adjectives or adverbs [20]. Author has employed PMI-IR (Point wise Mutual Information & Information Retrieval) algorithm to measure the similarity of pairs of words or phrases.

V. ADVANTAGES AND USAGE

- It mainly helps the concern to measure the overall performance of their business, especially on their online presence.
- It powers the business growth by providing actionable intelligence, instant involvement at any scale, economical intelligence and effective brand management.
- It allows to quickly identify the pitch of key conversations, issues and trends surrounding the business in real-time.
- Positive talking opinions around the brand can be identified to measure and evaluate digital strategy.
- Able to target the performance against the competitors and identify the changes in sentiments over time.
- Warnings can be set up for investors to be notified of positive or negative issues affecting the brand, which in turn helps making decisions to improve brand awareness and reputations.
- Allows the business to track both positive and negative reviews from high to low level view.
- Sentiment index can be built to measure company and product reputation.
- Helps to develop a strategy for responding to negative sentiment to positively influence opinion.
- Helps making strategic decisions by revising marketing style, customer service or product development.
- Finding those voices influencing customers and competitors.
- The emotion analysis helps in gathering new customer insights from social networks.

CONCLUSION

The paper has presented a surveyed analysis of the emotions of users and opinion mining to evident the approaches in E-business. The paper discussed the advantages of opinion mining for E-business. Emotions and opinion mining is a challenging task where the understanding of the problem and the solutions are limited. The industries need such approaches to know the customer and their views about the product and services. Many approaches and methods have been proposed but still it lacks the accuracy with higher gaps because of the language complexity. The needs and the technical challenges in the task will keep the field lively for the upcoming years.

Acknowledgments

The authors are thankful to the faculty members of the Department of Computer Science, Bharathiar University for their support. We also thank the friends and family for their inspiration and support.

References

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Nico H. Frijda, "Klaus Scherer's article on "What are emotions?" Comments", Social Science Information (SSI), SAGA Publications, Vol 46(3), pp. 381-443, 2007.

[2] Klaus R. Scherer, "Towards a dynamic theory of emotion", The component process of affective states, cognition and emotion, 1-1, 2001.

[3] R. Klaus Scherer, Angela Schorr, Tom Johnstone, "Appraisal Processes in Emotion : Theory, Methods, Research", published by Oxford University Press, 2001.

[4] Kim, S.-M., & Hovy, E., "Determining the sentiment of opinions", In Proceedings of the 20th international conference on computational linguistics, Geneva, Switzerland, pp. 1367, 2004.

[5] B. Pang, & L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts", In Proceedings of the 42nd annual meeting of the Association for Computational Linguistics, Barcelona, Spain, pp. 271–278, 2004.

[6] B. Pang, L. Lee, & Vaithyanathan, S., "Thumbs up? Sentiment classification using machine learning techniques", In Proceedings of the conference on empirical methods in natural language processing, Philadelphia, PA, USA, pp. 79–86, 2002.

[7] M. Setnes And H. Roubos, "Ga-Fuzzy Modeling And Classification: Complexity And Performance," Ieee Trans. On Fuzzy Systems, Vol. 8(5), Pp. 509–522, Oct 2000.

[8] H. Wu and J.M. Mendel, "Binary classification of ground vehicles based on the acoustic data using fuzzy logic rulebased classifiers," Tech. Rep. 356, USC-SIPI, Oct. 2002.

[9] J.A. Roubos, M. Setnes, and J. Abonyi, "Learning fuzzy classification rules from data," in RASC.

[10] A. Esuli, F. Sebastiani, "Determining the semantic orientation of terms through gloss analysis", Proceedings of CIKM , pp. 617–624, 2005.

[11] C. Strapparava, A. Valitutti, "WordNet-Affect: an affective extension of WordNet", Proceedings of LREC, vol. 4, pp. 1083–1086, 2004.

[12] B. Pang, L. Lee, S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques", Proceedings of EMNLP, pp. 79–86, 2002.

[13] J. Wiebe, E. Riloff, "Creating subjective and objective sentence classifiers from unannotated texts", Proceedings of the 6th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing), pp. 73–99, 2005.

[14] Renan Contreras, Oleg Starostenko, Vicente Alarcon-Aquino, Leticia Flores-Pulido, "Facial Feature Model for Emotion Recognition Using Fuzzy Reasoning", Advances in Pattern Recognition, Lecture Notes in Computer Science Volume 6256, pp 11-21, 2010.

[15] Ying Chen, Sophia Yat Mei Lee, Shoushan Li, Chu-Ren Huang,"Emotion Cause Detection with Linguistic Constructions", Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010), pages 179–187, Beijing, August 2010.

[16] E. Cambria, E. Hussain, C. Havasi, C. Eckl, "Affective space: blending common sense and affective knowledge to perform emotive reasoning", Proceedings of the 1st Workshop on Opinion Mining and Sentiment Analysis (WOMSA), 2009.

[17] G.Vinodhini and RM. Chandrasekaran, "Sentiment analysis and Opinion Mining: A survey", International Journal of advanced Research in Computer Science and Software Engineering vol. 2 Issue 6, 2012.

[18] Yi and Niblack, "Sentiment Mining in Web Fountain", Proceedings of 21st international Conference on Data Engineering, pp. 1073-1083, Washington DC, 2005.

[19] Christopher Scaffidi, Kevin Bierhoff, Eric Chang, Mikhael Felker, Herman Ng and Chun Jin "Red Opal: product-feature scoring from reviews", Proceedings of 8th ACM Conference on Electronic Commerce, pp. 182-191, New York, 2007.

[20] Peter D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," in Proc. 40th annual meeting of the ACL, pp. 417-424, July 2002.

[21] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.