

Examining Polarization of Emotions in Online Shopping Websites

Sanjeev Dhawan¹, Kulvinder Singh²

Faculty^{1,2}, Department of Computer Science & Engineering, University Institute of Engineering and Technology (U.I.E.T), Kurukshetra University, Kurukshetra, Haryana
Email: rsdhawan@rediffmail.com¹, kshanda@rediffmail.com²

Tanya Arora³

M.Tech.(Computer Engineering)³, Department of Computer Science & Engineering, University Institute of Engineering and Technology (U.I.E.T), Kurukshetra University, Kurukshetra, Haryana
Email: tanya.arora121@gmail.com³

Abstract

Today, everyone has exuberant ways to express his/her opinion, share content with others and access almost any information, hardly within a few seconds. The notoriety of online social networks and e-commerce websites has skyrocketed since last decade that they have become the indispensable need of the individual while conversing and shopping online. However, many users are harnessing these social and e-commerce websites as a mean of information and productivity. Online Shopping has overturned the idea people shop with so much ease and tremendous options. To understate insistent and arduous web surfing for comparison of various products, e-commerce websites have added a review and comment section for customers to facilitate social interaction among them which can be further analysed to detect emotion or opinion about a product -whether positive or negative. In this paper, a review of studies conducted on polarization of emotions using text mining is done to judge the shopping experience of the users and to increase the revenue for the online shopping websites from the users' weblogs.

Keywords –Emotion Mining, Information Extraction, Online Shopping, Pre-processing, Text Classification

1. Introduction

Online shopping has been surging enormously, keeping stride with the web. It facilitates customers with legion choices so that they are leveraged to buy from plethora of choices which are not available while traditional shopping. The vast number of items available on online shopping sites baffles customers to choose an item of their interest among similarly specified items which may be of different brands. Thus, it becomes cumbersome for customers to find unerring and befitting information regarding a specific product. To placate this trouble faced by customers, almost every e-commerce sites have

rendered a review and comment section for every product such that customers can post their views or opinions about the product that they have got delivered. This helps the buyer to get acknowledged about the product quality. Emotion or Opinion mining extends the probability to understand the users' comments and explain how a certain product or brand is comprehended- positively or negatively. But, data educed from these online review websites is unstructured and fuzzy in nature. Consequently, canvassing and evoking information patterns from such data sets are quite byzantine. Several researches have been carried out to examine

different ways of information extraction. The following discussed steps are conducted generally for analysing information- Pre-processing is to find the valuable data from raw available data and then assessing the valuable data so as to come to a peculiar conclusion. Two basic techniques of pre-processing- feature extraction and feature selection. Second step is Text classification which is the task of allocating unseen documents to suitable pre-defined categories. Various types of algorithms are used for text classification such as: k-nearest neighbour (kNN), Naïve Bayesian, Centroid based classifier, Support Vector Machine, and Rocchio classifier. The rest of the paper is structured as follows: section 2 analyses the previous work related to emotion or opinion mining and section 3 outlines the proposed work and section 4 at the end, concludes the paper and put up the future scope in emotion mining.

2. Related Work

'The beginning of wisdom is the definition of terms,' wrote Socrates. This instructive saying is highly applicable when it deals with the area of social media and e-commerce analysis, where any likeness of universal understanding on terminology is altogether lacking. According to Bang Po & Lillian Lee[1], Synonyms like opinion, view, belief, conviction, persuasion, and sentiment mean a verdict one grasps as right. Opinion means a decision thought out yet available for holdout. View implies a subjective opinion. Belief is an intended acceptance and rational proposal. Conviction holds to an unwaveringly maintained belief. Persuasion proposes a belief based on confidence. Sentiment is an opinion that reflects one's feelings. In presented paper, the reason behind information-gathering behaviour is to find out what other people think or perceive, reflects on their emotion or opinion i.e. whether positive or negative. With the abundance of opinion-rich resources such as online shopping sites and social networks, novel possibilities and disputes originate as people now can and do easily examine these websites to look for analysing the opinions of others and sentiments associated. In similar concern, Sheng Yu & SubhashKak [2] have found that if extracted and analysed properly, the data on social

media can lead to useful predictions of certain human related events. Most of the opinion mining methods attempted to detect the type of sentiment in texts: positive or negative. Mike Thelwalle *et al.* [3] expressed that texts frequently comprise of a combination of positive and negative sentiments and sometimes it is necessary to investigate both side by side. Barzilay & Lapata [4] introduced a data-driven method for learning the content-selection component for a concept-to-text generation system that determined which information should be the output of a natural language propagation system [4]. In text classification, k-nearest neighbour (KNN) is a simplest and efficient classifier (proposed by Cover P. & Hart T. [5]). However, due to its lazy learning approach (further surveyed by Sebastiani F. [6]) without pre-modelling, KNN is not cost efficient to categorize new documents when training set is large [6]. Rocchio classifier is also an established and mostly employed classifier for text classification (discovered by Joachims [7]), but when text data does not suit its implicit premise aptly, Rocchio classifier sustains. Duoqian Miao *et al.* [8] developed a hybrid system grounded on inconsistent precision rough set to aggregate the efficiency of both KNN and Rocchio techniques and overcome their weaknesses. Also, a reassessment of the KNN model was being approached by L. Cuccala *et al.* [9] as a statistical technique derived from a proper probabilistic model for carrying Bayesian inference on the arguments of the related model to appreciate the difficulties integral to it. A decision support model (DSM) was designed by Yue Dai *et al.* [10] that leverages various text mining methods, SWOT (Strength, Weakness, Opportunity, Threat) analysis to examine unstructured data from texts (e.g. newspapers, online sites, customer feedbacks and email) by using techniques like clustering, categorisation and sentiment analysis. The issue of automatically drawing out database values from the websites without learning instances was studied by Arvind & Hector [11] and then they defined the idea of a template and proposed a model that describes how values were encoded into pages using a template by presenting an extraction algorithm that used sets of words that had similar occurrence pattern in the input, to construct the template which

was then used to pull out values from pages [11]. Chunling Ma *et al.* [12] presented the approach for estimating emotions in natural-language texts which was a keyword spotting technique. Emotion investigation can be done by using either keywords spotting method or semantic network model. Huan Su ling *et al.* [13] examined those methods and achieved the better result with semantic network model. Yuki Murahmatsuet *al.* [14] built a program capable to perform entailment judgment on the account of word overlap, i.e. the similar types of the words in the texts. Similarly, any 'n' words within a document can have the same frequency, but one word might present the meaning of the sentence more accurately as compared to the others [15]. Consequently, Durga & Govardhana [15] found the new ontology based concept based model which analyses the terms on the sentences and documents level and effectively discriminates between non-important texts and important terms which help in understanding the correct sentence meaning [15]. To get to know how users are making choices from given information and what are the reasons that persuaded them, Lionel M. [16] presented different methods to detect influence in social media by emotion recognition. Jansen & Zhang [17] investigated the entire framework of these twitter posts, expressions, and the mobility in positive or negative sentiment and analysed the range, frequency, timing, and content of tweets. Narayanan *et al.* [18] observed sentiment analysis of conditional sentences. Due to the condition phrase, sentiments expressed in a conditional sentence can be hard to depict. This paper [18] aimed to determine whether opinions expressed on different topics in a conditional sentence are positive, negative or neutral. Thelwallet *al.* [5] also attempted to identify sentiments' polarity in text along with the strength of those sentiments. For this, he [19] created a new algorithm- SentiStength. Some research was also done by Pak & Paroubek [20] to introduce a method for an automatic collection of text documents that trains a sentiment classifier to determine polarity of documents. The classifier [20] is based on the polynomial Naive Bayes classifier that uses N-gram and Part of speech tags. Cvijikj & Michahelles [19] figured out the content shared on Facebook in terms

of topics, categories and shared sentiment for the domain of a sponsored Facebook brand page and displayed that Product, Sales and Brand were the three most discussed topics, while Requests and Suggestions, Expressing Affect and Sharing were the most common intentions for participation; to see the impact on social media marketing [19]. Consequently, from this background analysis, it can be inferred that different researchers have introduced different techniques by using distinct classifiers in their attempt to ameliorate accuracy while extracting emotional behaviour. However, to establish a time efficient and more accurate classifier has still remained the area of concern that needs improvisation.

3. Proposed Work

Most of the researchers exercised KNN method for text classification. But due to its inefficiency to provide much accuracy, an attempt will be made to customise or add new features to the existing basic KNN. It will definitely increase its accuracy in classifying pattern. Kernel-difference weighted KNN technique can also be adopted, which is further an extended version of weighted KNN. A feature of tagging can be applied after pre-processing which will figure out interesting texts that are recurring to save time.

4. Conclusion

In this paper, assorted researches have been surveyed down regarding polarization of emotions or opinions using distinct text classification methods after pre-processing of content available on e-commerce websites or any social media sites. By examining the reviews in review and comment section- either positive or negative about a product on any online shopping website, customers can aptly choose their item of interest in a nick of time and this opinion-affluent source of comments further helps us to classify which particular text of content is valuable to identify emotion behind that text i.e. information extraction. Techniques discussed can be enhanced further by customising classifiers- KNN, Naïve Bayesian, SVM or fuzzy classification by

adding new features so that text classification becomes more accurate and precise.

References

- [1] Pang B. & Lillian L., 'Opinion Mining and Sentiment Analysis', *Foundations and Trends in Information Retrieval* Vol. 2, Nos. 1–2, pp. 1–135, 2008.
- [2] Yu S. & Kak S., 'A Survey of Prediction Using Social Media', Springer, pp. 353-374, 2009.
- [3] Thelwall M., Buckley K., Paltoglou G. & Cai D., 'Sentiment Strength Detection in Short Informal Text', *Journal of the American Society for Information Science and Technology*, 61(12), 2544–2558, 2010.
- [4] Barzilay R. & Lapata M., 'Collective Content Selection for Concept-to-Text Generation', *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, pp. 331–338, Vancouver, October 2005.
- [5] Cover, T., & Hart, P., 'Nearest neighbour pattern classification', *IEEE Transaction on Information Theory*, 13(1), 21–27, 1967.
- [6] Sebastiani, F., 'Machine learning in automated text categorization', *ACM Computing Surveys*, 34(1), 1–47, 2002.
- [7] Joachims, T., 'A probabilistic analysis of the Rocchio algorithm with TFIDF for text categorization', In *Proceedings of the fourteenth international conference on machine learning*, 1997.
- [8] Miao D., Duan Q., Zhang H. & Jiao N., 'Rough set based hybrid algorithm for text classification', *Journal of Expert Systems with Applications*, 36(5), pp. 9168-9174, 2009.
- [9] Cucala L., Marin J., Robert C. & Titterton D., 'A Bayesian reassessment of nearest neighbour classification', arXiv:0802.1357v1 [stat.CO] 10 Feb 2008.
- [10] Dai Y., Kakkonen T. & Sutinen E., 'A decision-support model that combines text-mining technologies with two competitive intelligence analysis method', *International Journal of Computer Information System and Industrial Management Applications*, 3, pp. 165-173, 2011.
- [11] Arusu A. and Gracia-Molina H., 'Extracting structured data from web pages', *Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data*, New York, pp. 337-348, 2003.
- [12] Ma C., Helmut P. & Mitsuru I., 'Emotion Estimation and Reasoning based on Affective Textual Interaction', 3rd edition, Springer, 2005.
- [13] Ling H. S., Bali R. & Salam R., 'Emotion detection using keywords spotting and semantic network', In *Proceedings of International Conference on Computing and Informatics IEEE (ICOICI)*, Kuala Lumpur, pp. 1-5, 2006.
- [14] Muramatsu Y., Uda K. & Yamamoto K., 'Textual Entailment Recognition using Word Overlap, Mutual Information and Subpath Set', *Proceedings of the 2nd Workshop on Cognitive Aspects of the Lexicon (CogALex 2010)*, pages 18-27, Beijing, August 2010.
- [15] Durga A. K. & Govardhan A., 'Ontology based text categorization-telugu document', *International Journal of Scientific and Engineering Research*, 2(9), pp. 1-4, 2011.
- [16] Lionel Martin, 'User Behavior Under the Influence of Groups in Social Media', EDIC Research Proposal, 2009.
- [17] Jansen B. & Zhang M., 'Twitter Power: Tweets as Electronic Word of Mouth', *Journal Of The American Society For Information Science And Technology*, 60(11):2169–2188, 2009.
- [18] Narayanan R., Liu B. & Choudhary A., 'Sentiment Analysis of Conditional Sentences', *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 180–189, Singapore, 6-7 August 2009.
- [19] Cvijikj I. & Michahelles F., 'Understanding Social Media Marketing: A case study on Topics, Categories and Sentiment on a Facebook Brand Page', *MindTrek'11*, Tampere, Finland, September 28-30, 2011.
- [20] Pak A. & Paroubek P., 'Twitter as a corpus for Sentiment Analysis and Opinion mining', *Université de Paris-Sud, Laboratoire LIMSI-CNRS, Bâtiment 508, F-91405 Orsay Cedex, France*, 2012.