A Descriptive Study on Sentiment Analysis Measures and Methods

Meenakshi Madan

Research Scholar, Dept. of Computer Science & Engineering, Sat Priya Group of Institutions, Rohtak, Haryana, India

Rupali Malhotra

HOD, Dept. of Computer Science & Engineering, Sat Priya Group of Institutions, Rohtak, Haryana, India.

Abstract—Sentiment analysis is about to identify the hidden thoughts or emotions or the aspect based on detailed feature observation. Each of the verbal or verbal-expression taken from a user can be observed for sentiment identification or classification. In this paper, the basic process model for text based sentiment analysis is provided. The paper has described the standard process model of sentiment classification with relative process stages. The work behaviour of opinion mining and text processing is also provided in this paper. The paper also discussed the features, challenges, requirement and applications of sentiment recognition and classification.

Index Terms— Sentiment Analysis, Expression, Emotion, Opinion Mining.

I. Introduction

"What other people think" has always been an important piece of information for most of us during a decision making process. Internet has changed the way people express their views. Now-a-days people are more tech-friendly so internet is the best way to express their views and sentiments. Various companies are now turning to internet to collect reviews about their products.

Sentiment Analysis is a Natural Language Processing and Information Extraction task that aims to obtain writer's feelings expressed in positive or negative comments, questions and requests, by analyzing a large numbers of documents.

Sentiment Analysis is thus the process of processing the sentiments or feelings that are first gathered and then analysed. The most important part of gathering the information is to find out what other people think and then tracking their mood about a particular product. The information can be gathered through internet which is a huge repository of structured and unstructured data especially through the social media. Social media is generating a vast amount of sentiment rich data in the form of tweets, status updates, blog posts, comments etc.

The dependency on internet for analysing the sentiments has increased a lot in the past few years. Most people do online research and compare different options. When people wants to buy a new product, say a mobile phone they usually visit websites, watch videos and various blogs for product reviews and gather all the information about the product price, battery life or product quality before buying the actual product. This dependency has also made various companies to turn to internet to collect reviews about their products. The amount of user-generated data is too large for a normal person to analyse. So to automate this, various sentiment analysis techniques are used.

There are various feature-level extraction methods for collecting relevant features from text but the feature extraction is to be done in two phases to extract relevant features Firstly the raw reviews are processed to extract the stop list words. These are the words like- and, or, like, etc which does not impact the sentiment of a review so these are extracted to get a refined extracted words, after that data is processed to get the final review result based on a dataset of product reviews. Decision is based on the feedback of others. It's not only true for individuals it's also true for organizations. Here figure 1 is showing the basic sentiment extraction method with positive and negative sentiment class. Some example features are also shown in the figure.
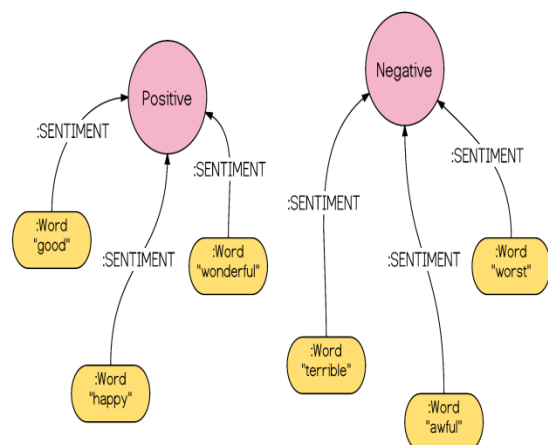

Figure 1 : Basic Sentiment Extration Method

Thus sentiment is an attitude, thought, or judgment prompted by feeling and Sentiment Analysis is the user generated data useful in knowing the opinion of the crowd after computation.

1.1 Text Mining/ Opinion Mining

Text-mining generally refers to the process of extracting interesting and non-trivial information and knowledge from unstructured text. Text mining encompasses several computer science disciplines with a strong orientation towards artificial intelligence in general, including but not limited to pattern recognition, neural networks, natural language processing, information retrieval and machine learning. Here figure 2 is showing the text processing method for sentiment extraction. The figure is showing the work stages.

Figure 2: Text Based Sentiment Classification Model

Opinion Mining is the knowledge extraction process applied on textual reviews to identify the hidden or expressed emotions or sentiments. It a computational form of natural language processing and having scope in all kind of web communication. This communication can be in the form of public reviews, user chat and the twitter posts. The observations can be formed to identify the sentences, phrases and words of these web posts. Some deterministic measures can be applied to take the featured decisions based on intended meaning of words or sentences.

1.2 Strategies Used For Sentiment Analysis

The. Symbolic or Knowledge base approach and Machine learning approach are the two strategies used for analyzing sentiments from the text. Symbolic approach requires a large database of predefined emotions and an efficient knowledge representation for identifying sentiments. Machine learning approach uses a training set to develop a sentiment classifier to classify sentiments.

1.3 Different Levels Of Analysis

Document Level : At this level , file contains group opinion and verify the complete file conveys a positive or negative sentiment. For example, given a product review, the system governs whether the review convey an overall positive or negative thoughts about the product.This task is commonly known as document-level sentiment classification.

Sentence Level : In Sentence Level, task goes to the take decision and determines whether each sentence conveyed a positive, negative, or neutral opinion. Neutral means no thoughts. Sentence Level is intently matched to emotional classification and peculiar sentences(called objective sentences)that express real information and conveys anomalous views and opinions.

Entity and Aspect Level : Aspect and entity level is not performing actually on the likes and dislikes. This level

performs like a fine tuner analysis. This aspect level also called as featured level. This always performs on the emotions(positive and negative) which are present at each opinion.

Here figure 3 is showing the work flow of sentiment analysis and classification method. The figure shows, as the textual review is accepted as input, a series of processes are applied to identify the sentiment class.
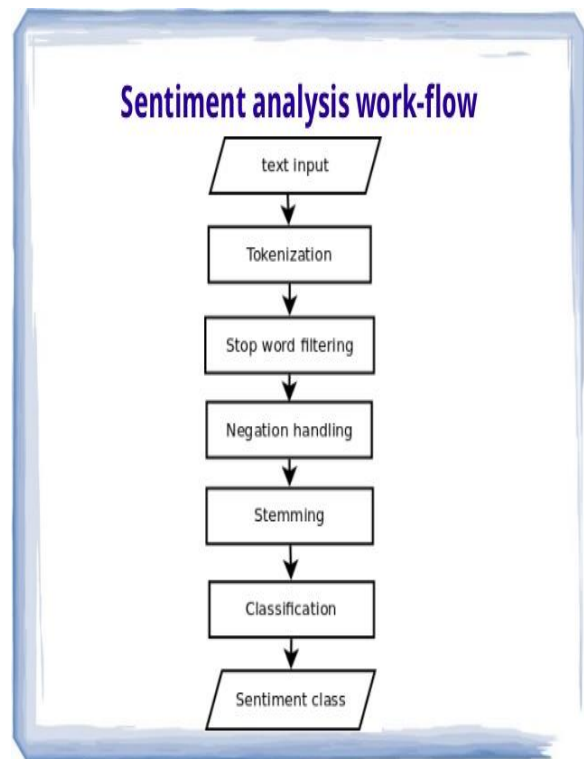

Figure 3 : Sentiment Analysis Work Flow

At first, the input text is divided in smaller works by separating it at the punchuation marks. As the words are identified, the stop list filteration method is applied over the text. After removing these words, the keywords are extracted from the input sentiment text. Later on the negotiation handling and stemming are applied to provide more specific keyword categorization. Finally, the relative classifier can be applied to identify the sentiment class.

II. EXISTING WORK

Kim Schouten(2015) provided a study work on aspect level analysis of sentiments over the reviews. Aspect-level sentiment analysis yields very finegrained sentiment information which can be useful for applications in various domains. Current solutions are categorized based on whether they provide a method for aspect detection, sentiment analysis, or both. Semantically-rich concept-centric aspect-level sentiment analysis is discussed and identified as one of the most

promising future research direction[1]. In the same year ,Wanxiang Che(2015) provided a sentence compression based sentiment analysis model. Different from the previous sentence compression model for common news sentences, Sent Comp seeks to remove the sentiment-unnecessary information for sentiment analysis, thereby compressing a complicated sentiment sentence into one that is shorter and easier to parse. Author apply a discriminative conditional random field model, with certain special features, to automatically compress sentiment sentences. The features proposed for Sent Comp, especially the potential semantic features, are useful for sentiment sentence compression[2].

Yahya Eru Cakra(2015) provided the work on stock price prediction using regression method for sentiment identification. The research is defined to predict the Indonesian stock market using simple sentiment analysis. Naïve Bayes and Random Forest algorithm are used to classify tweet to calculate sentiment regarding a company. The results of sentiment analysis are used to predict the company stock price. Author use linear regression method to build the prediction model. Presented experiment shows that prediction models using previous stock price and hybrid feature as predictor[3]. Mostafa Karamibekr(2012) has investigated the scope of sentiment analysis to cover various social issues. Author proposed an approach to take into account the role of verb as the most important term in expressing opinions regarding the social issues. Statistical and experimental results show that considering verbs not only is required and undeniable, but also improves the performance of sentiment analysis[4]. Neethu M S(2013) provided the work on machine learning techniques for sentiment analysis. Author analyzed the the twitter posts about electronic products like mobiles, laptops etc using Machine Learning approach. Author present a new feature vector for classifying the tweets as positive, negative and extract peoples' opinion about products[11]. Zhaoxia Wang(2014) has identified the social media issues and provided a solution using sentiment analysis method. Author survey the extant research literature on sentiment analysis and discuss various limitations of the existing analytical methods. Author propose a method that employs a new sentiment analysis scheme. Presented proposed method overcomes the limitations of the existing methods by not only improving the accuracy of the algorithm but also having the capability to perform analysis on non-English languages[12]

Siti Rohaidah Ahmad(2015) used a metaheuristic method for feature selection for sentiment analysis. This paper highlighted the comparative studies on the types of feature selection in sentiment analysis based on natural language processing and modern methods such as Genetic Algorithm and Rough Set Theory. The paper concluded that metaheuristic based algorithms have the potential to be implemented in sentiment analysis research and can produce an optimal subset of features by eliminating features that are irrelevant and redundant[5].

Mohsen Farhadloo(2013) defined a clustering score analysis based sentiment classification method. Author proposed a method to utilize bag of nouns instead of bag of words to improve the clustering results for aspect identification and a new feature set, score representation, that leads to more accurate sentiment identification. This scheme is based upon the three scores (positiveness, neutralness and negativeness) that are learned from the data for each term[6]. Xi Ouyang(2015) used the typical neural network modeling for sentiment analysis. Author propose a framework called Word2vec + Convolutional Neural Network (CNN). Author designed a suitable CNN architecture for the sentiment analysis task. Author use 3 pairs of convolutional layers and pooling layers in this architecture. Author also used the Parametric Rectified Linear Unit (PReLU), Normalization and Dropout technology to improve the accuracy and generalizability of Presented model[7].

Alpaslan Burak Eliacık(2015) has defined a user weight driven sentiment analysis model for twitter message classification. Author used the user metrics with Naive Bayes based sentiment analysis method and applies it to the finance field. Author also analyze the correlation between the mood of the financial community[8].

Charles B. Ward(2011) designed a framework for entity level sentiment analysis. Author propose Empath, a new framework for evaluating entity-level sentiment analysis. Empath leverages objective measurements of entities in various domains such as people, companies, countries, movies, and sports, to facilitate entity-level sentiment analysis and tracking. Author expect that Empath will encourage research that encompasses end-to-end pipelines to enable a large-scale text-driven monitoring and forecasting systems[9].

V.K. Singh(2013) provided a work on movie review based feature analysis method for improving the sentiment classification. Author has devised an aspect oriented scheme that analyses the textual reviews of a movie and assign it a sentiment label on each aspect. Author have also used Presented SentiWordNet scheme to compute the document-level sentiment for each movie reviewed and compared the results with results obtained using Alchemy API. The sentiment profile of a movie is also compared with the document-level sentiment result[10].

Wei Yen(2014) provided a NPL based method for sentiment analysis. The presented experiment consists of three main steps, which are subjectivity classification, semantic association, and polarity classification. The experiment utilizes sentiment lexicons by defining the grammatical relationship between sentiment lexicons and subject[13]. Addlight Mukwazvure(2015) has presented a hybrid model for news sentiment analysis. This research embarks on a hybrid approach to sentiment analysis of news comments which involves using sentiment lexicon for polarity detection (polarity will be classified as positive, negative and neutral)[15].

Ning Gu (2015) has proposed an interaction chain model for topic based sentiment analysis. Author organized messages as interaction-chains by taking advantages of the explicit interaction markers in microblog. Then, interaction-chains are clustered into different topics by comparing the similarity among them. After that, Author perform sentiment analysis using semantic-based SBV polarity algorithm. Author also proposed two heuristics according to the specificities of microblog[14] Umar Farooq(2015) used a word sense based feature adaptive method for sentiment analysis. Author proposed a feature level sentiment analysis system, which produces a summary of opinions about product features. A word sense disambiguation method is introduced which accurately determines the sense of a word within a context while determining the polarity. In addition, a heuristic based method is proposed in order to determine the text where opinion about a product feature is expressed[16].

III. CHALLENGES IN SENTIMENT ANALYSIS

There are several challenges in Sentiment Analysis :

Defining Opinion as positive or negative : First one is the word "opinion" itself which may be positive or negative depending on various situations. Here figure 4 is showing the spectrum of customer sentiment.
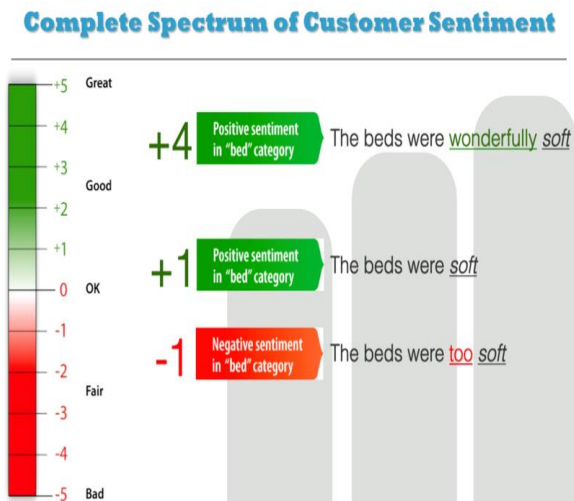

Figure 4: Spectrum of Customer Sentiment

Change in Feelings with Time : Second challenge is that the people feelings do change according to time. Sometimes they do like a product but after sometime their feelings about the same product change thus impacting the overall analysis.

Contradiction of Statements : Another challenge is that people can be contradictory in their statements. It means they have both positive and negative comments in the review which is sometimes manageable by analysing the sentence but sometimes it is not easy to analyse the exact view.

Differentiating Real Reviews : People in the virtual world comment without even analysing a product just by looking at previous reviews. So to differentiate between such reviews is almost impossible.

Missing Context : Another challenge is that the review collected from a social media may have a limit of characters like maximum limit of characters that are allowed in Twitter is 140 . So the context in which the user want to say anything may be missing.

Presence of Slang Words : Sentiment analysis through internet or social media is difficult compared to general sentiment analysis due to the presence of emotions, slang words and misspellings.

IV. APPLICATIONS OF SENTIMENT ANALYSIS

The applications for sentiment analysis are endless. More and more we're seeing it used in social media monitoring and voice of the customer (VOC) to track customer reviews, survey responses, competitors, etc. However, it is also practical for use in business analytics and situations in which text needs to be analyzed.

By computing customer satisfaction metrics, You can get an idea of how happy customers are with your products from the ratio of positive to negative tweets about them.
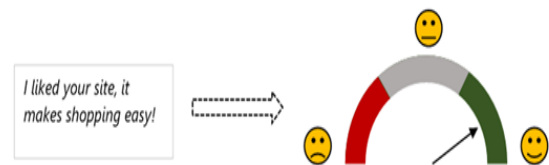

Figure 5 : Forcasting

SENTIMENT ANALYSIS APPLICATIONS
BUSINESS
Consumers voice
Brand reputation
Online advertising: Blogger Centric Contextual Advertising Dissatisfaction oriented online advertising
On-line commerce
POLITIC
Voting advise applications
Clarification of politicians' positions
PUBLIC ACTIONS
Real-world events monitoring
Legal matters "blawgs"
Policy or government-regulation proposals
Intelligent transportation systems

FINANCE
Prices of commodities and shares evolution
Financial risk individuation

Figure 6 : Applications of Sentiment Classification

We can forecast market movements based on news, blogs and social media sentiment. Another advantage is to identify the clients with negative sentiment in social media or news and to increase the margin for transactions with them for default protection. Here figure 5 is showing the forcasting applied based on the sentiment analysis.

It can also be used to find people who are happy with your products or services and their experiences can be used to promote your products. Here figure 6 is showing the various classes of sentiment class identification.

V. CONCLUSION

In this paper, a sentiment analysis for textual data processing is provided. The paper has discussed the challenges, application and scope of sentiment analysis. The paper defined the standard process model for sentiment extraction and classification with relative stage description.

REFERENCES

[1] K. Schouten and F. Frasincar, "Survey on Aspect-Level Sentiment Analysis," in IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 3, pp. 813-830, March 1 2016.

[2] W. Che, Y. Zhao, H. Guo, Z. Su and T. Liu, "Sentence Compression for Aspect-Based Sentiment Analysis," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 12, pp. 2111-2124, Dec. 2015.

[3] Y. E. Cakra and B. Distiawan Trisedya, "Stock price prediction using linear regression based on sentiment analysis," 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Depok, 2015, pp. 147-154.

[4] M. Karamibekr and A. A. Ghorbani, "Sentiment Analysis of Social Issues," Social Informatics (SocialInformatics), 2012 International Conference on, Lausanne, 2012, pp. 215-221.

[5] S. R. Ahmad, A. A. Bakar and M. R. Yaakub, "Metaheuristic algorithms for feature selection in sentiment analysis," Science and Information Conference (SAI), 2015, London, 2015, pp. 222-226.

[6] M. Farhadloo and E. Rolland, "Multi-Class Sentiment Analysis with Clustering and Score Representation," Data Mining Workshops (ICDMW), 2013 IEEE 13th International Conference on, Dallas, TX, 2013, pp. 904-912.

[7] X. Ouyang, P. Zhou, C. H. Li and L. Liu, "Sentiment Analysis Using Convolutional Neural Network," Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), 2015 IEEE International Conference on, Liverpool, 2015, pp. 2359-2364.

[8] A. B. Eliaçik and E. Erdoğan, "User-weighted sentiment analysis for financial community on Twitter," Innovations in Information Technology (IIT), 2015 11th International Conference on, Dubai, 2015, pp. 46-51.

[9] C. B. Ward, Y. Choi, S. Skiena and E. C. Xavier, "Empath: A framework for evaluating entity-level sentiment analysis," Emerging Technologies for a Smarter World (CEWIT), 2011 8th International Conference & Expo on, New York, NY, 2011, pp. 1-6.

[10] V. K. Singh, R. Piryani, A. Uddin and P. Waila, "Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification," Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), 2013 International Multi-Conference on, Kottayam, 2013, pp. 712-717.

[11] M. S. Neethu and R. Rajasree, "Sentiment analysis in twitter using machine learning techniques," Computing, Communications and Networking Technologies (ICCCNT),2013 Fourth International Conference on, Tiruchengode, 2013, pp. 1-5.

[12] Z. Wang, V. Joo, C. Tong and D. Chan, "Issues of Social Data Analytics with a New Method for Sentiment Analysis of Social Media Data," Cloud Computing Technology and Science (CloudCom), 2014 IEEE 6th International Conference on, Singapore, 2014, pp. 899-904.

[13] W. Y. Chong, B. Selvaretnam and L. K. Soon, "Natural Language Processing for Sentiment Analysis: An Exploratory Analysis on Tweets," Artificial Intelligence with Applications in Engineering and Technology (ICAIET), 2014 4th International Conference on, Kota Kinabalu, 2014, pp. 212-217.

[14] N. Gu, D. y. Sun, B. Li and Z. Li, "Sentiment Analysis for Topics based on Interaction Chain Model," Intelligence and Security Informatics Conference (EISIC), 2015 European, Manchester, 2015, pp. 133-136.

[15] A. Mukwazvure and K. P. Supreethi, "A hybrid approach to sentiment analysis of news comments," Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), 2015 4th International Conference on, Noida, 2015, pp. 1-6.

[16] U. Farooq, T. P. Dhamala, A. Nongaillard, Y. Ouzrout and M. A. Qadir, "A word sense disambiguation method for feature level sentiment analysis," 2015 9th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), Kathmandu, 2015, pp. 1-8.