DIFFERENT APPROACHES TOWARDS SENTIMENT ANALYSIS

Geeta Bisht
Research Scholar, Department of Computer Science and Technology, Jayoti Vidyapeeth Women's University, Jaipur, E-mail: Geeta13sbisht@gmail.com

Professor Dr. Shobha Lal
Department of Science and Technology, Jayoti Vidyapeeth Women's University, Jaipur,Rajasthan, India, E-mail: dean.fet@juvw.ac.in

Abstract: Sentiment Analysis is characterized as the method of data mining, view, study, also sentence to predict the emotion of the phrase through the natural language Processing (NLP). The sentiment analysis requires the division of the text into "+ve", "-ve" or "Neutral" three stages. It examines the details and marks the 'better' and 'worse' emotions as good and bad. Thus, the World Wide Web (WWW) has been a major repository of personalized or user-generated raw data in recent years. In WWW using social media and e-commerce sites like Facebook, Twitter, Amazon, Flipkart etc. millions of people share their thoughts, their emotions and perceptions about things. This increases raw data on internet that is an incredibly high source of information, either positive or negative, for any decision-making process. The science of emotion analysis has tended to process such enormous data automatically. The primary objective of SA is to define and characterize the data's polarity on the Network. Sentiment analysis is text-based analysis, but the exact polarity of the sentence is challenging to find.

Keywords: Sentiment Analysis, techniques, emotion, machine learning, supervised learning.

Introduction: In the machine survey of emotions, perceptions & feelings expressed in the text [1]. Sentiment analysis attempts to identify unique content present in various references and assess an author's mind-set against a subject or a text's general disposition. Wiebe et al. [2] defined subjectivity as the textual depiction of another's feelings, opinions, attitudes, decisions, beliefs, 3 and speculations. The terms vision, sentiment, opinion, and belief are interchangeably used, but slightly vary [3]. • Opinion: A conclusion (each authority appeared to have a distinct opinion) thought out and open to debate. • Viewpoint: Subjective view. • Faith: thoughtful consent & intellectual assent • Sentiment: relaxed perspective expressing one's emotions (her activist feelings are wellknown).SA is conducted on Site substance created by users that include thoughts, emotions, or views. The product analysis, a discussion post, a blog, or a tweet which assesses an item may be an opinionated text. For goods, challenges, entities, organizations, or a program, for example, the views indicated may be about something or anyone.

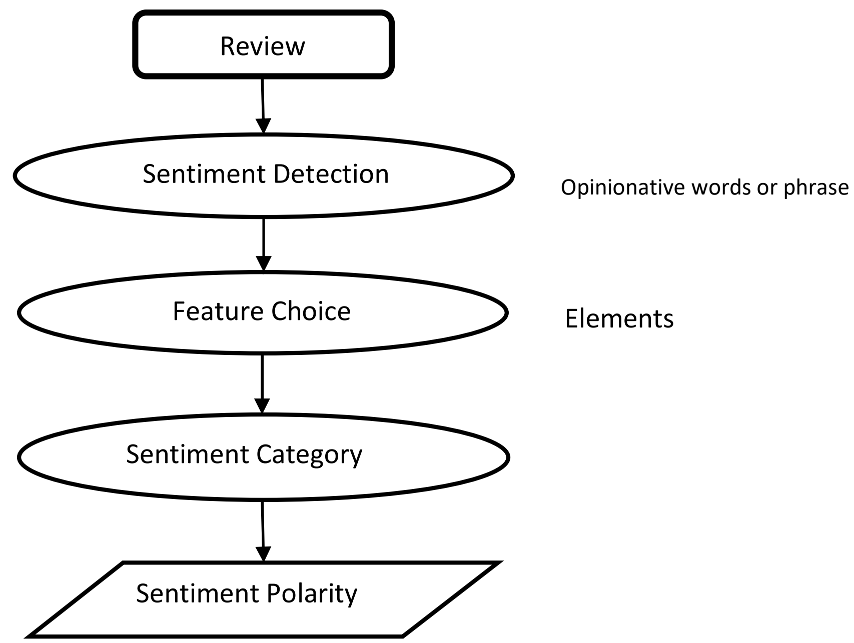

Fig.1.1. Sentiment Analysis process on product

Sentiment Analysis Types: There are four kinds of SA

1) Fine-grained Sentiment Analysis requires assessing the polarity of an individual's viewpoint. It may be a straightforward linear positive/negative emotion distinction. Depending on the uses case, this form will also go into the higher specification (for example, very good, positive, neutral, bad, very negative) (for example, as in five-star Amazon reviews).

2) Emotion detection is applied to classify signs of specific emotional states submitted in the text. Typically, there is a mixture of lexicons & ML algorithms which decide what is what & why.

3) Aspect-based sentiment analysis aims to figure out what people think about a certain part of the product. Remember the brightness of the smartphone's flashlight. Aspect-based analysis is widely preferred in product analytics to keep track of how a product is viewed by consumers & what its strengths & disadvantages are.

4) Intent Analysis is all about the case. Its objective is to detect what kind of objective is expressed in the text. It is often utilized in customer support systems to make simpler the workflow.

Levels of analysis: SA contains 3 levels:

1. Document level- In this SA organizes all whole text opinion into variant sentiment, to a goods & service. Document level categorizes opinion text into a +ve, -ve, or neutral sentiment. For making simpler the mission, it is believed that each text‘s whole opinion is totally held by a particular opinion holder & is about a individual item. Numerous ML methods happen for this mission. This conducts to sentence level SA.

2. Sentence level –In this SA defines whether every group of words conveys a +ve, -ve, or impartial opinion, for a goods or provision. This level is used to evaluate & remarks that include a sentence & write down with the operator [5]. At this level, research has been done on recognition of subjective sentences in a document from a combination of purpose & subjective sentences & later, the sentiment initiation of these subjective sentences is decided.

3. Entity & Aspect level –In this level opinion mining & explanation related to the operation. The categorization affects by detecting & obtaining goods includes by the cause data. This level is utilized the require sentiments regarding needed include in an assessment.

Techniques of Sentiment Analysis: For sentiment analysis, there are 2 major techniques: ML related & lexicon related. Some study experiments are related these two approaches to achieve comparatively improved results as well.

1. **Machine learning based techniques**: The methodology of ML related to SA is primarily part of supervised classification. The methodology based on Machine Learning in which two sets of documents are mandatory: training & a set of tests. An automated classifier uses a training set to learn the distinctive features of texts, & a examine set is utilized for verify how good the classifier works. In order to identify the feedback, a variety of machine learning approaches were introduced. In sentiment analysis, ML practices for example- Naive Bayes (NB), max entropy (ME), & support vector machines (SVM) have attained considerable popularity continues with the compilation of datasets for handling. Training a classifier on the training data is the next step. When a supervised classification technique is chosen, the selection of features is an essential choice to make. They will teach us how they portray papers.
   - Categorization is achieved by evaluating the characteristics of a given text with sentiment lexicons in an unsupervised system whose sentiment values are calculated previous for their usage. The emotion lexicon includes lists of terms & phrases utilized for connect the emotional thoughts and perceptions of individuals. For example, examine the text for which emotion needs to be 8 identified, beginning with +ve and -ve word lexicons. Then it is +ve if the text contains more positive word lexicons, otherwise it is -ve. The Sentiment Analysis lexicon-based strategies are unsupervised learning, so prior preparation is not necessary to identify the results.
   - POS data: Part of Speech is used to explain the meaning that is used in order to direct the collection of features. A marker that reflects its position/role in the grammatical sense will be applied to each word in phrases in POS tagging. For e.g, we can classify adjectives and adverbs with POS tags that are normally used as indicators of sentiment.
   - Contradictions: Contradiction is also an essential aspect which is believed as it has the possibility of changing an opinion.
   - Opinion words & phrases: Opinion words & phrases are words & slogans which convey +ve & -ve sentiments. The major methods to classify the semantic positioning of an opinion word are numerical related to lexicon. Hu & Liu et al. use WordNet to confirm whether the obtained procedural has a +ve / -ve polarization.

2. **Lexicon related method:** Categorization is achieved by evaluating the characteristics in a provided text by sentiment lexicons in an unsupervised system in which sentiment ideals are calculated preceding to their use. The emotion lexicon includes records of terms & phrases utilized to communicate the emotional thoughts and perceptions of individuals. For example, examine the text for which emotion needs to be identified, beginning with +ve & -ve word lexicons.

It is +ve if the text contains more positive word lexicons, or else it is -ve. The Sentiment Analysis lexicon-based strategies are unsupervised learning, so prior preparation is not necessary to identify the results. The fundamental actions of the lexicon related to the methods are summarized below as:

a) Modify overall sentiment score of the text $s \leftarrow 0$. 3. Tokenize script. In this every token, examined if it is current in a sentiment vocabulary. If gesture is current in glossary,

(i) If token is +ve, so $s \leftarrow s + w$.

(ii) If token is -ve, so $s \leftarrow s - w$. See the overall content sentiment achieves,

b) If s > threshold, then categorize content like +ve.

c) If s < threshold, so it categorizes the text as -ve.

3. **Hybrid Techniques:** Some testing methods have shown that the mixture of both ML & lexicon-based techniques increases the efficiency of emotion categorization. Mudinas et al. proposes the method of conception-stage emotion analysis, pSenti, that is established with integrating methods focused on 9 lexicons and learning. By using a lexicon/learning symbiosis, the key benefit of their hybrid approach is to accomplish the best of all worlds-stability, legibility from a carefully constructed lexicon, & high precision from an efficient supervised learning algorithm. For initial sentiment identification, their framework utilizes a sentiment lexicon constructed by applying social tools. The sentiment lexicon presently contains of 7048 sentiment, comprising words by wildcards, & in the range of − 3 to + 3 sentiment values are identified.

Outcomes of experiments demonstrate that while a common-purpose sentiment lexicon offers only slight enhancement in precision, it leads to more substantial improvement when integrating domain-specific dictionaries. A 2-step group was accomplished by the system. A classifier is available in step 1 to forecast the camera function being addressed. A classifier is prepared in phase 2 to calculate the sentiment correlated with the feature of the camera. In last stage, the effects of both the stages of forecast are mixed together to harvest the last forecast. The lexicon data is combined into conservative SVM learning in both steps. They reached a polarity correctness of 66.8 percent.

Features of Sentiment Analysis:

i. **Term Presence vs. Term Frequency**: In conventional Information Retrieval & Text Classification activities, term frequency has always been thought vital. However, Pang-Lee et al. (2002) discovered that term presence is more essential than term frequency in SA. i.e. binary-valued feature vectors with binary values in which the items simply signify whether a word exists (value 1) or not (value 0). It is not counter instinctive, considering the lots of cases where even a single string emotion carrying word will reverse the polarity of the whole sentence. The existence of uncommon terms includes more details than the occurrence of regularly occurring words, a condition known as Hapax Legomena.

ii. **Term Position**: Words that occur in particular areas in the document have greater sentiment or significance than words happening in a different place. i.e. equivalent to IR, where words in subject descriptions, subtitles, and abstracts, among other areas, are given greater weight than those in the body. Since the text includes +ve terms all through, the existence of a -ve sentiment at the final sentence 10 defines the sentiment in the example given in Section 1.3.c. As a consequence, terms that occur in the first limited sentences & the last only some sentences of a text are offered greater significance than words appearing elsewhere.

iii. **N-gram Features**: This is commonly used in Natural Language Processing tasks because they can extract meaning to some extent. It's arguable if greater order n-grams are helpful. When it comes to identifying movie reviews by emotion polarity, Pang et al. (2002) find that unigrams outperform bigrams, but Dave et al. (2003) discovered that bigrams & trigrams do best in certain circumstances.

iv. **Subsequence Kernels:** The majority of SA works use a word or sentence level model, with the effects averaged over each words/sentences/n-grams to provide consistent model productivity for every review. Subsequences are used by Bikel et al. (2007). The perception is that the function space indirectly seized by subsequence kernels is wide enough to remove the requirement for direct information engineering & also emotion modelling at the word or sentence stage.

Conclusion: Sentiment-based analysis is a rapidly developing discipline with a wide range of applications. Despite the fact that estimation-based analysis projects are tough to complete due to their regular language processing origins, due to their prominence in recent years, significant progress has been accomplished. There has been a lot of study done in this topic, however there are still a lot of challenges, such as sentiment analysis forming unstructured content. Although word reference-based methodology takes less processing time than regulated learning approach however, accuracy isn't enough for decision-making. Supervised learning approach gives better outcomes. The work shows a picture of the research status in the field of sentiment analysis in which different strategies are considered for effective assessment. This work will help analysts to have a clear picture of the different points of view about the sentiment analysis and the difficulties associated with its implementation.

REFERENCES:

[1] Lee, Ching-Pang, et al. Electrochemical machining process for forming surface roughness elements on a gas turbine shroud. No. US 6379528. General Electric Company (Schenectady, NY), 2002

[2] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up? Sentiment classification using machine learning techniques." arXiv preprint cs/0205070 (2002).

[3] Bikel, Daniel M., and Jeffrey Sorensen. "If we want your opinion." International conference on semantic computing (ICSC 2007). IEEE, 2007.

[4] Dave, Kushal, Steve Lawrence, and David M. Pennock. "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews." Proceedings of the 12th international conference on World Wide Web. 2003.
[5] AlDayel, Abeer, and Walid Magdy. "Stance Detection on Social Media: State of the Art and Trends." arXiv preprint arXiv:2006.03644 (2020).
[6] Hu, Minqing, and Bing Liu. "Mining and summarizing customer reviews." Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. 2004.
[7] Annett, Michelle, and Grzegorz Kondrak. "A comparison of sentiment analysis techniques: Polarizing movie blogs." Conference of the Canadian Society for Computational Studies of Intelligence. Springer, Berlin, Heidelberg, 2008
[8] Morin, Frederic, and Yoshua Bengio. "Hierarchical probabilistic neural network language model." Aistats. Vol. 5. 2005..
[9] Grimes, Andrea, and Richard Harper. "Celebratory technology: new directions for food research in HCI." Proceedings of the SIGCHI conference on human factors in computing systems. 2008.
[10] Bekmamedova, Nargiza, and Graeme Shanks. "Social media analytics and business value: a theoretical framework and case study." 2014 47th Hawaii international conference on system sciences. IEEE, 2014.
[11] Kleindienst, Sara, et al. "Chemical dispersants can suppress the activity of natural oildegrading microorganisms." Proceedings of the National Academy of Sciences 112.48 (2015): 14900-14905.