Exploration of Sentiment Analysis Using Feature-Based Classification, and Addressing Negations Using NLP

Tripti Anand[1], Rekha[2]

[1]M. Tech. Scholar, [2]Assistant Professor
CBS Group of Institutions, Jhajjar (Haryana)
Maharishi Dayanand University, Rohtak

***Abstract*: Sentiment analysis is techniques of text analysis which detect text polarity automatically. Sentiment research is often called impression mining and is one of the NLP's (Natural Language Processing) main activities. Study of the emotions has attracted tremendous popularity in recent years. Individuals are supposed to build a mechanism capable of defining and classifying thoughts or emotions as expressed in an online document. Consumers frequently face the trade-off in purchasing choices and nowadays if you choose to purchase a consumer good you choose customer feedback and conversation regarding the product in online forums on the internet. When taking their purchasing choices, often customers use feedback shared by other users. People have a propensity to speak out about various institutions. Opinion mining has grown in popularity as a result. Sentiment Analysis assesses how this viewpoint shared regarding the object has a favorable or a negative orientation. Consumers ought to determine which subset of knowledge they choose to use. The method by which contextual meaning is defined and derived from raw data is known as sentiment analysis. An effective method for predicting feelings may enable one to collect views from the internet and predict the tastes of online consumers, which could prove useful for economic or marketing research. There are so far few specific problems that predominate in this research community, namely, classification of emotions, feature-based classification, and addressing negations. This paper provides a study of the approaches and strategies that exist in the area of sentiment analysis, and obstacles.**

***Keywords*: Sentiment Analysis, Sentiment Research, Natural Language Processing, Negative Orientation**

I. Introduction

Sentiment Studies are the applications of human language, text interpretation and computerized linguistics to classify and derive contextual knowledge from source sources, and are often related to under Opinion Mining. In certain terms, Sentiment analysis is a sort of linguistic method to measure the public's sentiment on a certain commodity or topic. It accepts an input list and gives a sensory level (very negative to very positive) in the range [-1 t0 1]. This includes the development of a method for collecting and evaluating feedback on the commodity provided in blog articles, reports, comments or tweets. Analysis of emotions in different forms may be helpful. For example, it may help to impede the effectiveness of a marketing strategy or introduction of new goods, decide which variations of a product or service become common, and also evaluate whether populations have the same or different characteristics. Sentiment analyzes rely on behaviors, whereas conventional research is based on empirical study.

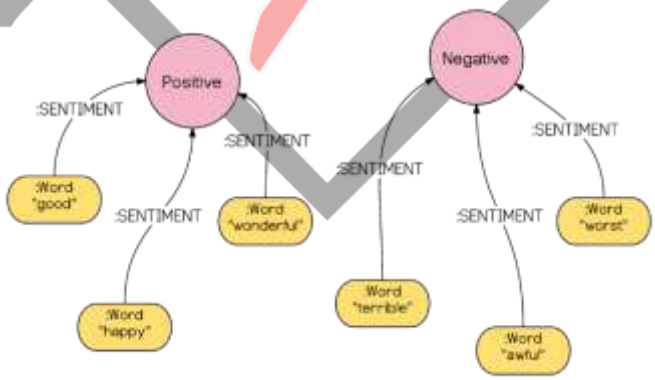


Fig: Sentiment Analysis

For emotion analysis there are three key areas of research: description of thinking, description of feelings depending on the features, and overview of opinions. Sentiment grouping includes the description of the views of other items in entire papers. Feature-based Grouping Sentiment on the other hand takes into consideration views on the characteristics of other products. The overview opinion function is not the conventional text review, because even the software functions where consumers share their opinions are compromised.

Natural language processing (NLP) is an area in computer science, artificial comprehension and computer linguistics that deals with machine-to - human languages and, in particular, programming computers to effectively manage broad corporates in natural

languages. Natural language processing tasks also include interpreting the language, developing natural languages (sometimes from structured functional structures, computer-readable), linking language to the experience of the user, handling or integrating human-computing communication systems.

Social networking has now developed into a common forum for citizens to share their opinions to the world. The Web is quickly evolving from a static platform to an immersive media. Consumers today will obtain details and produce contents effectively. Media, sites, journals, etc. They are the primary outlets of knowledge to the media. This electronic word of mouton offers a modern and observable knowledge base for several uses and it is called feel research. The document includes all cases and opinions, and may be analyzed through natural language therapy to achieve a specific amount of viewpoint opinions. This is the way to define and derive contextual details from raw data. Sometimes named viewpoint mining is sentiment-analyze. Sentiment research not only allows the customer to access more and more information with a single mouse click on specific goods and services, but also lets him make a more educated decision. A emotions report may be recorded because the emotion is synthesized as optimistic, negative or neutral in the entire text. This may be a paragraph on the grounds of which particular phrases of emotions are listed in the document. SA may be a statement dependent on the polarity of the words in a paragraph. Yes, it is optimistic or bad environment to classify relational interpretation role opinions conveyed in a letter. Natural language processing (NLP) is an area in linguistics, covering computers, artificial intelligence, the person (natural) and linguistic experiences. Of starters, in the analysis of the product, it decides if the characteristics of the product discussed by the reviser are good, negative or neutral. The statement, for example, "The lifespan of the mobile's batteries is too short," communicates a view on the mobile object's "existence of its batteries" and is pessimistic. By fact, the quality of an emotion analysis method is how good it is aligned with individual views. It is commonly calculated by consistency and warning.

II. Sentiment Classification

Classifications of emotions are based on the optimistic, negative or neutral polarity. Which ensures the views may be defined as optimistic, negative or neutral. In addition, a fourth form is a positive opinion that accepts recommendations for enhancement of the product. The literature survey performed applies to two kinds of approaches, namely computer testing and textual orientation, in relation to sentiment analysis. The grouping of emotions presumes that the opinion paper communicates views about a particular person or topic. Opinions provide knowledge that can be separated into two specific categories: truth, normally factual claims regarding a particular person (objet) or occurrence and opinions that are emotional in nature and that are the opinion holders' opinions regarding the entity. For decision-making, all evidence and views are useful.

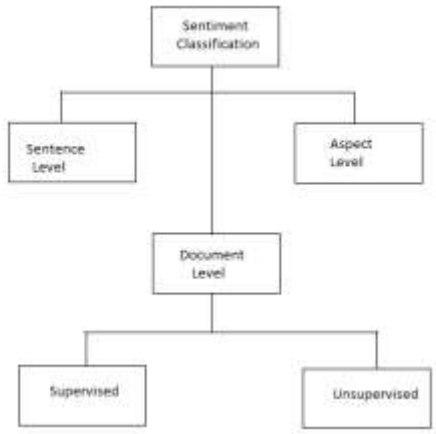


Fig: Sentiment Classification

Therefore, natural language methods of processing (NLP), particularly in the identification of document sentiment, are used in this area. At the crossroads of NLP and content processing, modern emotion identification thus shares a range of characteristics with other functions, knowledge extraction, text mining, computational linguistics, psychology, predicate research, etc.

Opinions can also be divided into three types: primary views, contrasting views and indirect views. Opinion holders strike the goal directly. Indirect beliefs are inferred either as in languages or as opposite as in sarcasm. In contrast, opinion leaders typically differentiate all individuals. Meaning analysis is always carried out in these studies at one of three levels: the stage of a text, sentence or attribute.

Document Level

The analysis of text opinion attempts at classifying the entire paper as optimistic or bad. Either of the two recognition systems is a lot of hard research. Is a controlled process and an unattended method for creating a feeling of document stage?

Supervised method:

Classification of emotions is carried out at record stage. Sorting attitude may be used to work with four types of optimistic, negative, neutral and supportive issues as controlled sorting. In addition, supervised requests to create relationships between articulated

opinions and text parts for machine-learning algorithms including SVM Support Vector Machines. Many scientists have considered controlled learning strategies in the SVM and Naïve Bayes to be effective.A satisfactory selection of apps is one of the most critical activities of emotion classification. Some main characteristics are:

Terms and Frequency: These functions are N-grams associated with independent terms (characters). The most common features are often included in the grouping of typical theme documents.

Speech section: Any term can also be relevant as a speech component (POS). The term for each part of speech (POS) may vary. For example, adjectives provide many details about the feel of text.

Words of sentiment and phrase: terms of emotion or thoughts are expressions in a language used to convey optimistic or negative feelings. Great, great, sweet words are optimistic emotions, and terrible, mediocre and dangerous words are harmful emotions, for example.

Rules of opinion: There are also other words and formulations in the vocabulary that may often be used to convey or suggest thoughts and views, aside from emotions or sentences...

Sentiment shifters: They are terms used for changing the emotions, from good to bad, to optimistic or supportive. Sentiment shifters:

The most critical form of mood changes are offensive terms. The expression "This mobile does not like me," for example, is a negative. Syntactical dependency: Words reliance-based sorting or dependency tree characteristics. Researchers are still seeking to do that.

Unsupervised method:
Uncontrolled categorization at sentence level is done. Two classifications are unattended, focused on lexicon and syntactic trends. The grouping of sentences and elements of emotions on the basis of lexicons can be used. The predominant markers for emotion recognition are terms and sentences with thought or thinking. This should then be very normal to use spontaneous thinking focused on terms and phrases.

Turney demonstrated an uncontrolled estimation of learning for the interpretation of a reform as planned or not implied. He learned whether the terms are good or bad and how powerful the appraisal is by using a good seed word ("excellent") and a bad seed word ("poor") to convey the wise shared knowledge (PMI). The Word's semantic orientation was named this meaning. The technique analyzed the semanthematic orientation in these phrases, and added semanthese orientation of all the phrases in order to measure a summary orientation via an audit to identify expressions which fit those grammatical characteristics (descriptive terms and intensifiers). He achieved 74% precision for a corpus of object ratings.

Sentence Level
Their task is to define each sentence of the text as a positive or negative opinion at a particular point in emotion classification. The nostalgic study of sentence standard described the polarity. This point is similar to paper point, but every word fulfills it here. Nevertheless, repetitive sentences in the document cannot support the quality of the sentence. There are two steps of the standard sentence emotion of each paragraph: first, the polarity of the subjective paragraph is inferred, each sentence being marked as either subjective or objective.

Subjectivity analysis also is called the function of classifying a statement as subjective or objective. Including the subsequent descriptive phrases, which are considered a ranking on the emotion stage, are defined as voicing positive or negative opinions. The polarity of each phrase is determined in the sentence level emotion study. Close to a study of emotions at record stage but done at sentence point.

This means that every sentence includes an opinion on an object and a certain dimension and that certain sentences cannot be opined (target). The descriptive sentences include terms of judgment that help to decide the feeling towards the object. A two-step distinction is rendered for each word; first, each term is marked as arbitrary or factual, and then each phrase's polarity is implied. The views will often include nuanced phrases. For these situations, the designation of penalty rates is not effective.

Aspect Level
This means that there are multiple individuals and facets of a text. Classification of aspects and classes includes the exploration of each of these things, aspects and emotions.

Naive Bayes procedure
This is a Bayesian theorem-based classification strategy, provided that the predictor variables are autonomous. Generally speaking, including a particular function in a class is not linked to any other function in the class. The Naive Bayes classification indicates that the fruit may be called an apple. It can be red, round or an appetizer about 3 inches in diameter. All of these traits individually bring the possibility that the fruit is an apple, because these traits are interdependent or dependent on certain traits. Therefore, it is considered "naive". Naive Bayesian models are easy to create and are especially useful for large amounts of data. In addition to consistency, naive Bayes classification methods are considered more complex.

Likelihood Class Prior Probability

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)}$$

Posterior Probability Predictor Prior Probability

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

• P(c) is the probability after the predictor (x, attribute) specified by the category (c, target).
• The probability of the previous class is P(c).

• P(x) is the probability of the predictor in the specified category.

• P(x) is the predictor probability of the predictor variable.

Why does the algorithm of Naive Bayes work?
Let's get it by way of a case. Below I've got a training environment data collection and the associated 'Play' goal variable (suggested practice options). Now we will decide whether or not teams are practicing on the environment level. Take the instructions below to do so.

Phase 1: Convert configuration data into frequency table
Step 2: Create probability table by calculating probability of cloudy sky = 0.29 and probability of playing game 0.64.

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency Table		
Weather	No	Yes
Overcast		4
Rainy	3	2
Sunny	2	3
Grand Total	5	9

Likelihood table				
Weather	No	Yes		
Overcast		4	=4/14	0.29
Rainy	3	2	=5/14	0.36
Sunny	2	3	=5/14	0.36
All	5	9		
	=5/14	=9/14		
	0.36	0.64		

Step 3: Compute the retrospective probability using the Bayesian Naive equation for each category. Forecast results are the most likely category in history.

Naive Bayes (Naive Bayes) uses a general method to estimate the probability of a group depending on certain attributes. This method is mainly used for text classification and some types of problems.

III. Bayes Procedures

• **Echtzeit Vorhersage:** Naive Bayes is an impatient classification and should be simple. It can also be used to present real-time forecasts.

• **Multiclass prediction:** This multiclass prediction algorithm is also well established. Here you can estimate the likelihood of several target variable groups.

• Text Classification/Spam Filtering/Sentiment Analysis: Compared to other algorithms, Naive Bayes classifier, which is mainly used for text classification, has higher success rate due to its excellent results in multi-class problem and the law of freedom. Become. Therefore, it is commonly used in the processing of fraud (identification of fraudulent emails) and sensory analysis (distinguishing between good and negative customer sentiment for social network data).

• Collaborative Filtering and Naive Bayes Classification Model produces a Recommendation Framework that utilizes computer and data mining techniques to extract input from consumers and determine whether they want to use a tool.
How is the Naive Bayes in Python a simple model built?

Similarly, scikit programming (Python library) helps create a naive Bayesian foundation for Python. Under the scikit learning library, there are three types of naive Bayes.

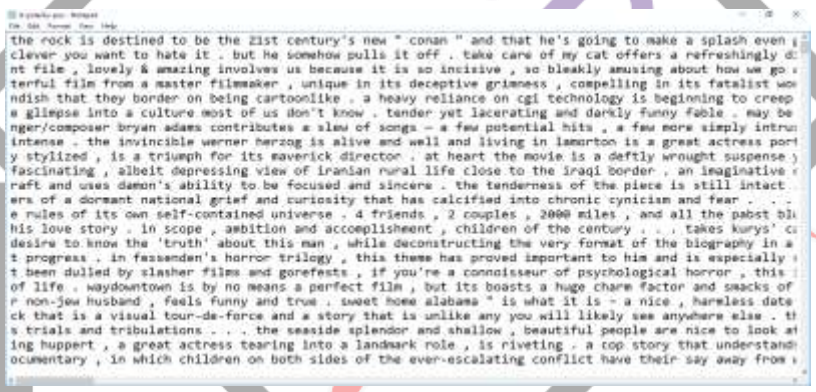
• **Gaussian:** Used to label the natural distribution of elements.

• **Multinomial:** used for discrete counting purposes. We've got a text classification problem, for instance. There we will see Bernoulli trials that are one phase forward so, instead of "term in the document" it should be called that "the amount of times the outcome x i were found in the n trials" were "how many term in the document happens."

• **Bernoulli:** When the vjectors are binary (i.e. zeros and zeros), the binomial model is helpful. Another use is the classification of the text using the 'word bag pattern,' in which the 1 and the 0s are 'words in the document' and 'words do not appear in the document'

Review Data Negative

Review Data Positive



IV. Conclusion

The study of perception or opinion mining is a research area, in which thoughts, perceptions or sentiments towards other individuals are studied. The paper discusses a significant issue in the study of emotions, the categorization of emotion polarity. Sentiment research continues to be enhanced and progressed still. In fact, a complicated term poses other problems, such as the polarity. Furthermore, the natural language vocabulary is all that causes difficulties. This thesis highlights the basic concepts of emotion analysis.

V. Future Scope

• This paper provides a study of sensation research approaches and processes and problems on the field.
• Classification of emotion, Classification of technology, instruments for the study of feelings present.
• A successful approach to predict emotions will help us to derive views from the web and predict online tastes of our consumers, which may be useful for economic or marketing analysis.

References

[1] T. Nasukawa, "Sentiment Analysis: Capturing Favorability Using Natural Language Processing Definition of Sentiment Expressions," pp. 70–77, 2003.
[2] K. Dave, I. Way, S. Lawrence, and D. M. Pennock, "Mining the Peanut Gallery□: Opinion Extraction and Semantic Classification of Product Reviews," 2003.
[3] X. Ding, S. M. Street, B. Liu, S. M. Street, P. S. Yu, and S. M. Street, "A Holistic Lexicon-Based Approach to Opinion Mining," pp. 231–239, 2008.

[4] E. Marrese-Taylor, J. D. Velasquez, and F. Bravo-Marquez, “Opinion Zoom: A Modular Tool to Explore Tourism Opinions on the Web,” 2013 IEEE/WIC/ACM Int. Jt. Conf. Web Intell. Intell. Agent Technol., pp. 261–264, Nov. 2013.
[5] J. Zhu, H. Wang, M. Zhu, B. Tsou and Matthew M, “Aspect-Based Opinion Polling from Customer Reviews", " Ieee Transaction On Affective Computing”, vol. 2, NO. 1, January-March 2011.
[6] E. Haddi, X. Liu, and Y. Shi, “The Role of Text Pre-processing in Sentiment Analysis,” Procedia Comput. Sci., vol. 17, pp. 26–32, Jan. 2013.
[7] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet Allocation,” vol. 3, pp. 993–1022, 2003.
[8] T. Hofmann. P. latent, “ semantic indexing,” In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in inFmnation retrieval, SIGIR '99, pages 50-57, New York, NY, USA, 1999. ACM
[9] R. Moraes, J. F. Valiati, and W. P. Gavião Neto, “Document-level sentiment classification: An empirical comparison between SVM and ANN,” Expert Syst. Appl., vol. 40, no. 2, pp. 621–633, Feb. 2013.
[10] R. Arora and S. Srinivasa, “A Faceted Characterization of the Opinion Mining Landscape,” pp. 1–6, 2014.