SENTIMENT ANALYSIS - A CONCLUSIVE TOOL FOR OPINION MINING, IMPLEMENTING MACHINE LEARNING

Ms. Chris Snow don[1] Dr.P.Tamil Selvan[2]

ABSTRACT

Sentiment Analysis extracts the inner feeling behind a text, graphic, audio or video file. In the present scenario all sectors from different parts of the globe generate uncontrollable amount of data every day. Sentiment analysis senses the subject and polarity of diversified data and presents them to a structured and organized output. The main aim of this article is to discuss the areas of application and polarity classification in sentimental analysis as well as the major tools and algorithms commonly used to figure out the outcome.

Keywords: *NLP, Sentiment Analysis, polarity classification, lexicon, machine learning.*

1. INTRODUCTION

Opinion mining works on Natural Language Processing that helps to identify the opinion or sentiment behind a text. Sentiment analysis usually extracts polarity, subject and opinion owner from the selected dataset. More than 80% of data input we are getting nowadays are in unstructured format. Opinion mining helps to give a structure to these data and helps to classify them under various criteria.

Opinion can be subdivided into facts and opinions. Facts are always objective while opinions are subjective. Opinions can be direct or comparative as well as implicit and explicit. The top layer of sentiment analysis acts as a two-way classifier which classifies a text under subjectivity classification and polarity classification. Subjectivity classification mainly identifies the subject behind the opinion whereas polarity classification marks the expressed opinion as positive, negative or neutral. The scalability and real time analysis that can be done with sentimental analysis make it an efficient tool for structuring data.

2. STAGES IN POLARITY CLASSIFICATION

Polarity classification is the fundamental phase in sentiment analysis. This classification can be done at document, sentence and sub-sentence level. The first step in polarity classification is to identify the dataset. Datasets can be a blog, a social media like Facebook or twitter, customer review pages, etc.

The second phase is the preprocessing stage, were the data or subject is extracted effectively for better classification. The main NLP techniques applied at this stage are:

Stemming: It is the process of removing the suffices and finding the root word.

Tokenization: Given a sequence of characters or a document, this NLP procedure throws away all the unwanted details and breaks down the whole document into meaningful tokens.

Speech Tagging: Converting each sentence into a set of tuples, usually in the form word or tag.

Table 1: Part of Speech and Tag

Speech Tagging

{ INCLUDEPICTURE "https://media.geeksforgeeks.org/wp-content/uploads/pos-1.jpg" * MERGEFORMATINET}

Parsing: Arranging words by understanding the relationship among words.

[1]Research Scholar, Department of Computer Science, Karpagam Academy of Higher Education

[2]Assistant Professor, Department of CS, CA & IT, Karpagam Academy of Higher Education

3. COMMON TECHNIQUES FOR CLASSIFICATION

Sentiment or opinion classification techniques that are widely in use now are represented in Fig. 1.
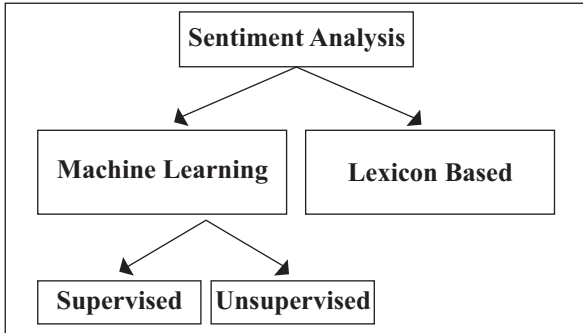
Fig. 1: Sentiment Analysis techniques

Machine learning techniques can be divided into supervised and unsupervised learning. Supervised learning method predicts the output based on various input conditions. The algorithms used in the supervisory method contain labeled data. Though it is time-consuming, it will produce a more accurate output. The process flow is illustrated in fig. 2.

> {   INCLUDEPICTURE
> "https://www.guru99.com/images/1/030819_1026
> _SupervisedM2.png" * MERGEFORMATINET
> }

Fig. 2: Supervised Learning

There are various supervised learning methods like decision tree and rule based, probabilistic and linear classifiers.

In unsupervised learning there are no prewritten procedures from input to output. The system as a whole learns through observation. It learns from the present data and classifies the input based on this learning. This classification of similar data can be done with the help of classifier algorithms like hierarchical clustering, clustering nearest neighbors clustering, etc.

Lexicon-based method uses a predefined set of words for determining polarity. For this method a template of positive and negative words is required with a sentiment

value attached with each word. A combining mathematical function is applied after assigning numeric sentiment values to the words in the dataset. This will give a final prediction on the overall sentiment of the dataset. Flowchart of the lexicon-based approach is shown in fig. 3.

> {   INCLUDEPICTURE
> "https://www.researchgate.net/profile/Eman_Younis/
> publication/272463313/figure/fig3/AS:29497837176
> 8332@1447339732334/Lexicon-based-sentiment-
> analysis-approach.png" * MERGEFORMATINET}

Fig. 3: Lexicon Method

Dictionary-based and corpus-based are two classifications in lexicon-based approach. Dictionary-based method extracts opinion-based words and search for - similar words in the predefined pooled set of words.

Corpus is a large body of text, which includes a bit of semantic information as well. In corpus-based approach a data-driven method is adopted where we will have access not only to sentiment labels, but also to a meaningful context.

Table 2: Corpus based sentiment Analysis

Annotated Corpus	
Sentence	**Polarity**
The review is good	**Pos**
It is damaged	**Neg**
The mobile is blue	**Neu**

4. Conclusion

Sentiment analysis is an area where there is scope for more and more innovations and developments. The data of an individual have become the nuclear part of all marketing strategies. In this paper, different methods for sentiment classification have been discussed for classifying the polarity of data. Lexicon-method is gaining more popularity as it seems to be more flexible and dynamic. Rather than classifying the subject as positive and negative

outcomes, we can bring more clarity to the opinion by categorizing the sentiments into marketing, opinion, complaint, suggestion, query, news, etc. Contextual semantic search can be performed to attain this. Algorithms implementing precision, recall, F-measure using deep learning techniques can be deployed for getting more accuracy.

5. REFERENCES

[1] Jason Yip"Algorithmic Trading using sentiment analysis on News Articles"Towards Datascience, 2018

[2]"https://towardsdatascience.com/@shashankgupta_54342?source=post_page-----6c94d6f58c17 sentiment analysis: concept analysis and applications", towards datascience, 2018

[3] Xiang Fang, Justin Zhan" Sentiment Analysis using product review data", Journal of Big Data, 2015

[4] Soonh Taj" Sentiment Analysis of News Articles"A Lexicon based approach", 2nd international conference on computing Mathematics and Engineering Technologies, 2019

[5] Vishal Shersat, Rajkumar S Jagadaleand Sachin deshmukh" Document level sentiment analysis from news articles", International conference on computing control and automation, 2017

[6] Wataru Souma, Irene Vodenska,Hidaki Aoyana" Enhanced news sentimentanalysis using Deep Learning methods", Journal of computational social science, 2018

[7]Artem Opperman"Sentiment Analysis with Deep Learning of Netflix reviews", Towards data science, 2019

[8]Sahar Sohangir, Dingding wang, Anna pomernets, Taghi M Khoshoftaar" Big data: Deep Learning for financial sentiment analysis" Journal of Big Data, 2018