International Journal of Engineering and Techniques - Volume 3 Issue 3, May-June 2017

RESEARCH ARTICLE OPEN ACCESS

Sentiment Analysis in Natural Language Processing

[1]Neha Gaur (M.Tech Student), [2]Dr. Neetu Sharma (HOD)

[1,2]Ganga Institute of Technology and Management, kablana (Jhajjar)

Abstract:

Sentiment analysis is type of analysis techniques which analysis text that automatically detect polarity of text. Sentiment analysis also called as opinion mining which is one of the major tasks of NLP (Natural Language Processing). Sentiment analysis has much popular in recent years.People are intended to develop a system that can identify and classify opinion or sentiment as represented in an electronic text. Consumers regularly face the trade-off in purchase decisions so nowadays if one wants to buy a consumer product one prefer user reviews and discussion in public forums on web about the product. Many consumers use reviews posted by other consumers before making their purchase decisions. People have a tendency to express their opinion on various entities. As a result opinion mining has gained importance. Sentiment Analysis deals with evaluating whether this expressed opinion about the entity has a positive or a negative orientation. Consumers need to decide what subset of available information to use.The process of identifying and extracting subjective information from raw data is known as sentiment analysis. An accurate method for predicting sentiments could enable us, to extract opinions from the internet and predict online customer's preferences, which could prove valuable for economic or marketing research. Till now, there are few different problems predominating in this research community, namely, sentiment classification, feature based classification and handling negations. This paper presents a survey covering the techniques and methods in sentiment analysis and challenges appear in the field.

1. INTRODUCTION

Sentiment analysis (also known as opinion mining) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. In other words we can say Sentiment analysis is a type of natural language processing for tracking the mood of the public about a particular product or topic. Its major task is Identify and extract sentiment in given string . It takes an input string and assigns a sentiment rating in the range [-1 t0 1] (very negative to very positive).It involves in building a system to collect and examine opinions about the product made in blog posts, comments, reviews or tweets. Sentiment analysis can be useful in several ways. For example, in marketing it helps in judging the success of an ad campaign or new product launch, determine which versions of a product or service are popular and even identify which demographics like or dislike particular features.

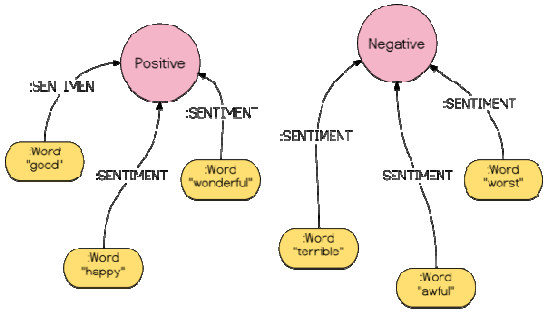

Fig 1 : Sentiment Analysis

Sentiment analysis concentrates on attitudes, whereas traditional text mining focuses on the analysis of facts. There are few main fields of research predominate in Sentiment analysis: sentiment classification, feature based Sentiment classification and opinion summarization. Sentiment classification deals with classifying entire documents according to the opinions towards certain objects. Feature-based Sentiment classification on the other hand considers the opinions on features of certain objects. Opinion summarization task is different from traditional text summarization because only the features of the product are mined on which the customers have expressed their opinions. Opinion

summarization does not summarize the reviews by selecting a subset or rewrite some of the original sentences from the reviews to capture the main points as in the classic text summarization. Sentiment Analysis uses various classification techniques to identify the tone of a given piece of text. It indicates whether the text is positive, negative or neutral. This analysis can be aggregated over large sets of data and the resulting information can be helpful in different contexts.For example, in the sentence, "The life of the battery of this mobile is too compressed", the opinion is on "life of the battery" of the mobile object (target) and the opinion is negative. Many day to day life applications require this level of detailed analysis because in order to make product upgrade one needs to know what components and/or features of the product are liked and disliked by consumers. Such information has not come across by sentiment and subjectivity classification. Natural language processing (NLP) computer science, Artificial intelligence, and computers and human (natural) concerned with interactions between languages is an area of Linguistics. For instance, in a product review, it identifies features of the product that have been commented on by the reviewer and determines whether the comments are positive, negative or neutral. SA can be phrase based where the phrases in a sentence are classified according to polarity. In fact, to identify the emotion analysis task views expressed in a text is positive or negative weather.

2. OBJECTIVE

The main objective of this research is that people may know the what exactly sentiment are kept behind the review written by a person .In this we are using the real world movie review. Because is we are using good data our result is also effective. This paper presents a survey covering the techniques and methods in sentiment analysis and challenges appear in the field. Its main objective to extract opinions from the internet and predict online customer's preferences, which could prove valuable for economic or marketing research. Sentiment analysis is text analysis techniques that automatically detect the polarity of text.

For example, in marketing it helps in judging the success of an ad campaign or new product launch, determine which versions of a product or service are popular and even identify which demographics like or dislike particular features. Sentiment analysis is also called as opinion mining. Sentiment analysis not only helps in allowing the user to get more and relevant information about different products and services on a mouse click, but also helps in arriving at a more informed decision.

3. PROBLEM SPECTIFICATION

Many consumers use reviews posted by other consumers before making their purchase decisions. People have a tendency to express their opinion on various entities. As a result opinion mining has gained importance. Sentiment Analysis deals with evaluating whether this expressed opinion about the entity has a positive or a negative orientation. Consumers need to decide what subset of available information to use. Sentiment Analysis face number of problems.

- Namely
- Sentiment Classification
- Featured based classification
- Negation

3. SENTIMENT CLASSIFICATION

Sentiment classifications are based on polarity, which can be positive, negative, or neutral. That's mean opinions may be classified into positive, negative, or neutral. Opinionated documents contain information which can be broadly categorised in two categories: facts and opinion. Both facts and opinions are useful in decision making. sentiment analysis is often conducted at one of the three levels: the document level, sentence level, aspect level.

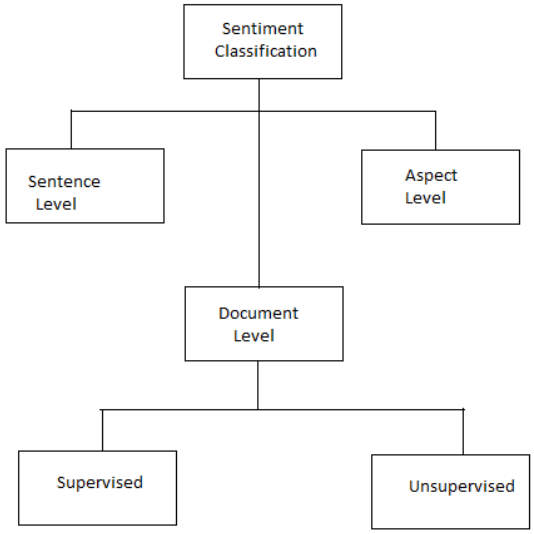

Fig : 2 Sentiment Classification

Document Level

Document level sentiment classification aims to classify the entire document as positive or negative. There is much actual work use one of the two types

of classification techniques which are a Supervised method and Unsupervised method to build level document sentiment.

Sentence Level

It is one level of sentiment classification its work is to determine each sentence in the document as positive or negative opinions. Sentence level sentiment analysis has classified the polarity.

Aspect Level

It supposes that a document has a hold opinion on many entities and their aspects. Aspect level classification needs discovery of these entities, aspects, and sentiments for each of them.

4. SOFTWARE REQUIREMENTS

4.1 PYTHONE Python 3.6.1 is now the latest maintenance release of Python 3.6 and supersedes 3.6.0. Get 3.6.1 here.Python 3.6.0 is the newest major release of the Python language, and it contains many new features and optimizations. See the What's New In Python 3.6 document for more information.Many organizations are using Python these days to perform major tasks. You don't necessarily hear about them because organizations are usually reserved about giving out their trade secrets. However, Python is still there making a big difference in the way organizations work and toward keeping the bottom line from bottoming out.

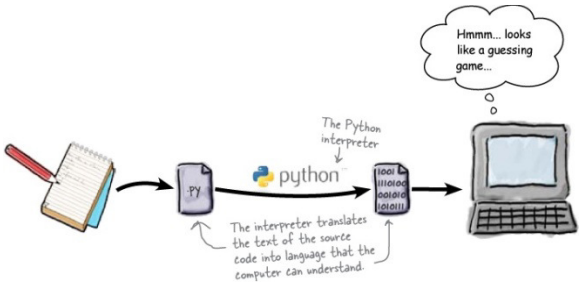

Fig : 3 Python process

Major new features of the 3.6 series, compared to 3.5

- Among the new major new features in Python 3.6 are:
- Preserving Keyword Argument Order
- Simpler customization of class creation
- Local Time Disambiguation
- Literal String Formatting
- Adding A Secrets Module To The Standard Library
- Add a private version to dict
- Underscores in Numeric Literals
- Adding a file system path protocol
- Preserving Class Attribute Definition Order
- Adding a frame evaluation API to CPython
- Make os.urandom() blocking on Linux (during system startup)
- Asynchronous Generators (provisional)
- Syntax for Variable Annotations (provisional)
- Change Windows console encoding to UTF-8
- Change Windows filesystem encoding to UTF-8
- Asynchronous Comprehensions

5. TECHNOLOGY USED

5.1 NLTK

NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active discussion forum.

Thanks to a hands-on guide introducing programming fundamentals alongside topics in computational linguistics, plus comprehensive API documentation, NLTK is suitable for linguists, engineers, students, educators, researchers, and industry users alike. NLTK is available for Windows, Mac OS X, and Linux. Best of all, NLTK is a free, open source, community-driven project.

NLTK has been called "a wonderful tool for teaching, and working in, computational linguistics using Python," and "an amazing library to play with natural language."

Natural Language Processing with Python provides a practical introduction to programming for language processing. Written by the creators of NLTK, it guides the reader through the fundamentals of writing Python programs, working with corpora, categorizing text, analyzing linguistic structure, and more.

NLTK is the most famous Python Natural Language Processing Toolkit, here I will give a detail tutorial about NLTK. This is the first article in a series where

I will write everything about NLTK with Python, especially about text mining and text analysis online

NLTK is intended to support research and teaching in NLP or closely related areas, including empirical linguistics, cognitive science, artificial intelligence, information retrieval, and machine learning.

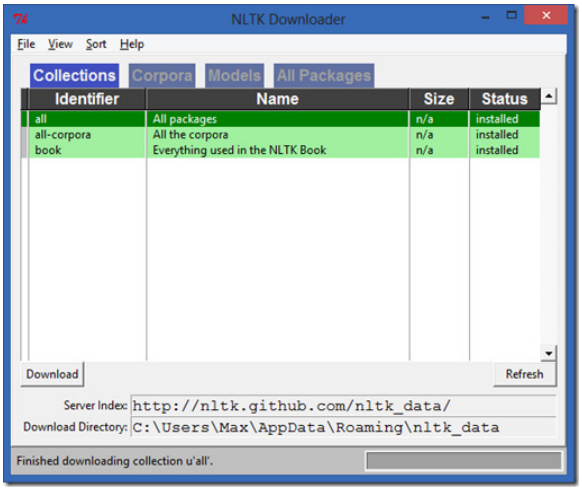

Fig : 4 NLTK Downloader

5.2 Naive Bayes algorithm

It is a classification technique based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. For example, a fruit may be considered to be an apple if it is red, round, and about 3 inches in diameter. Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability that this fruit is an apple and that is why it is known as 'Naive'.Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

5.3 Test Data And Trained Data

In many areas of information science, finding predictive relationships from data is a very important task. Initial discovery of relationships is usually done with a training set while a test set and validation set are used for evaluating whether the discovered relationships hold. More formally, a training set is a set of data used to discover potentially predictive relationships. A test set is a set of data used to assess the strength and utility of a predictive relationship.

Test and training sets are used in intelligent systems, machinelearning, genetic programming and statistics.

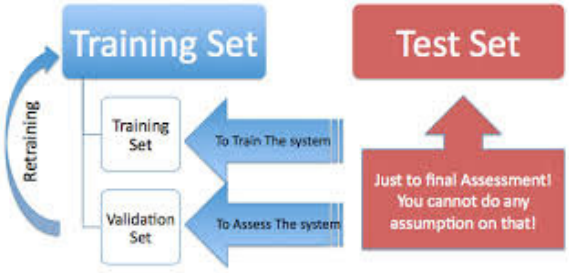

Fig : 5 Test data And Trained Data

6. METHODOLOGY

Data collection : The data which is used in this paper is a set of movie reviews collected from http://www.cs.cornell.edu/people/pabo/movie-review-data/rt-polaritydata.README.1.0.txt

Lower case converter :

In this set of review first convert from upper case to lower case reason behind this is that one word count once not again. For example – 'SMALL' . 'small 'it count both at once mean SMALL = small.

Reduce the short word :

In this we reduce the word which having the length less than 3 because these words not make a sense of sentiment .For example –the, is etc.

Feature vector formation :

Sentiment tokens and sentiment scores are information extracted from the original dataset. They are also known as features, which will be used for sentiment categorization. In order to train the classifiers, each entry of training data needs to be transformed to a vector that contains those features, namely a feature vector. For the sentence-level (review-level) categorization, a feature vector is formed based on a sentence (review). One challenge is to control each vector's dimensional it.

7. PROCESS OF WORK

1. In this first we collect the data or review from the data source.
2. After that data is collected first convert whole data into lower case .
3. And reduce the all stop words in data with the help of NLTK for this you need to download NLTK
4. And reduce all words which have length less than 3.
5. After this we use navie bayes classifier which provide.
6. Generate the result in three forms precision, Accuracy , recall .
7. For this vectorization process you need to download python version because this contain python which contain the code.
8. So you need to provide the platform where this .py extension file may be run.

8.CONCLUSION

Sentiment analysis or opinion mining is a field of study that analyzes people's sentiments, attitudes, or emotions towards certain entities. This paper tackles a fundamental problem of sentiment analysis, sentiment polarity categorization. Sentiment Analysis still need to improve and progress. Moreover, there are many challenges like the polarity in a complex sentence. In addition, the vocabulary of natural languages is a lot which causes difficulty. This survey highlights the basic ideas about Sentiment Analysis and then explains in details the Sentiment Classification, Technique Classification, tools that available for Sentiment Analysis, and a new feature which is Product Aspect Ranking.

9.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.