Sentiment Analysis Using Data Dictionary

[1]Shivani Rohilla, [2]Mukesh Rawat,[3]Anmol Jain

[1,2] Deptt. of CSE, MIET, Meerut

[3]M.Tech CSE, MIET, Meerut

Abstract

With the advances in technology about sentiment analysis and predictive analytics, it has opened many avenues for researchers and enterprises to understand human mental state better. The proposed challenge is to know the emotion/mood of a person, to help in eliminating any negative state of mind that might have adverse effect on his/her daily life. This paper introduces techniques for the emotional analysis of a person in textual form which can be further used to predict the behavior of a person. Sentiment analysis is a standalone Java application that provides user with options to analyze the feeling of a person in different domains. Unlike others, our proposed methodology will provide the user to analyze a document by providing the phrase or messages in it. We are making use of a tool named facepager which is used for fetching the data from facebook. Facepager takes a bunch of facebook page names, and then enables the user to gather all posts or likes as well as a number of other types of data; these can be exported as data files from data dictionary for further processing and analysis. Author proposes a methodology that can further be used by psychiatrists in the treatment of a patient by seeing his insight and hence it can be very beneficial for them.

Keywords: Sentiment analysis, data dictionary, facepager

1. Introduction

Sentiment analysis is an application of natural language processing which can be used in text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to voice of the customer materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine. Basically, sentiment analysis aims to determine the attitude of a speaker, writer, or other subject with respect to some topic or the overall contextual polarity or emotional reaction to a document, interaction, or event. The attitude may be a judgment or evaluation, affective state (that is to say, the emotional state of the author or speaker), or the intended emotional communication.
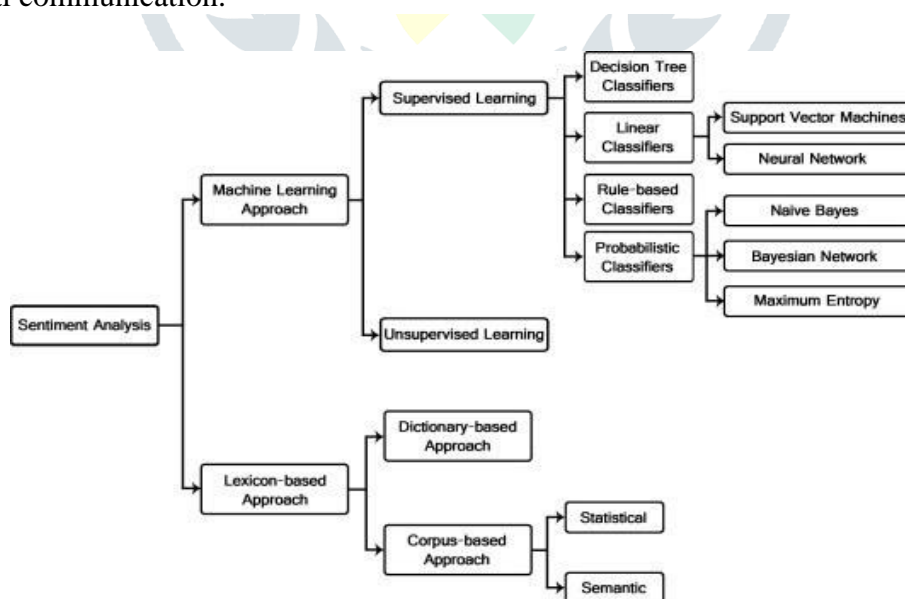

Fig1: Methods to achieve sentiment analysis

1.1 Need of sentiment analysis

Sentiment analysis can be a useful tool for an organization to understand what its employees are feeling, and take remedial measures to salvage the development areas, while leveraging the strengths. The applications of sentiment analysis are broad and powerful. The ability to extract insights from social data is a practice that is being widely adopted by organizations across the world.

1.2 Natural Language Processing

Natural-language processing (NLP) is an area of computer science and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to fruitfully process large amounts of natural language data. NLP aims to combine artificial intelligence technology with human interactions beautifully.
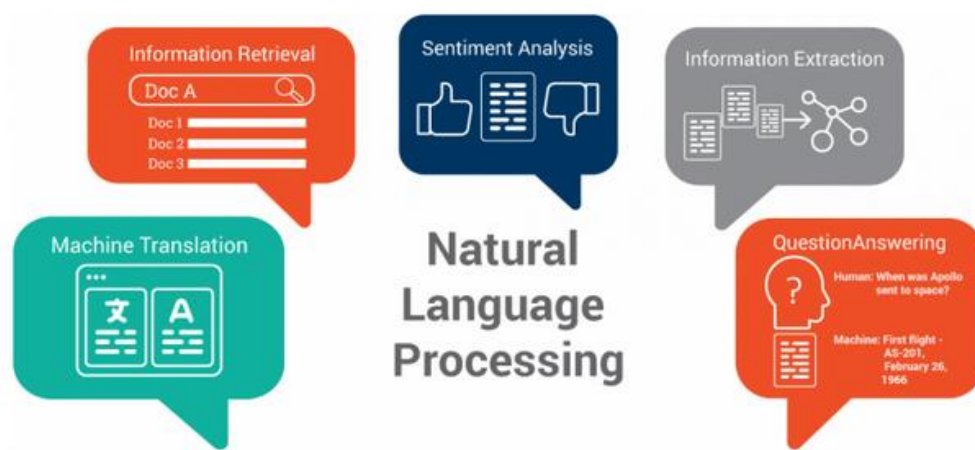

Fig2: Use of natural language processing

NLP techniques incorporate a variety of methods to enable a machine to understand what's being said or written in human communication—not just words—in a comprehensive way. The potential use cases of NLP can be remarkable. Some of most common and popular uses include sentiment analysis, keyword interests.

1.3 Constitutes of a good data dictionary

Data dictionary is a set of information describing the contents[3], format, and structure of a database and the relationship between its elements, used to control access to and manipulation of the database. A crucial data set for any kind of text mining is a dictionary. As for sentiment analysis there are two big families of analysis algorithm. Both leverage dictionaries. The lexicon-based approach is where the ultimate score is calculated based on a per-word score from the dictionary and machine learning approach, where dictionaries are used to reduce data dimensionality. A good dictionary for our purpose contains words bearing a strong positive or negative connotation, but does not contain neutral words. It may include two- or three-word phrases to improve handling negations and other common cases, like "can't stand." Ideally, the dictionary should come from the same topic and writing style that will be analyzed. (A lexicon for short texts like tweets written by teenagers will be quite different from a lexicon for diplomatic messages.)[10] So, the best dictionary for our case would be based on tweets about movies, and if we want to get sophisticated we may need it will need to go beyond words and short phrases and include the role of a word in a sentence. For that level of model we would need a dictionary that would include features such as the part of speech for each word.

Using Sentiment analysis one doesn't have to read complete message to detect the feelings of an individual. Also, the sentiment analysis reduces mental work of psychiatrists as well as normal human being. Using Data dictionary, one can detect persons' emotions expressed in the text. Author proposes a technology which is very easy to use as compared to previous technologies used for analysis of emotions.

2. Proposed System

The system describes how to summarize the text and analyze the text given in to it. This desktop application provides user with options to analyze the feeling of a person in different domains. The purpose of this paper is to present a detailed description of sentimental analysis using online data dictionary. It will explain the purpose and features of the software, the interfaces of the software, what the software will do, the constraints under which it must operate and how the software will react to external stimuli. This paper is intended for both the users and the developers of the system.
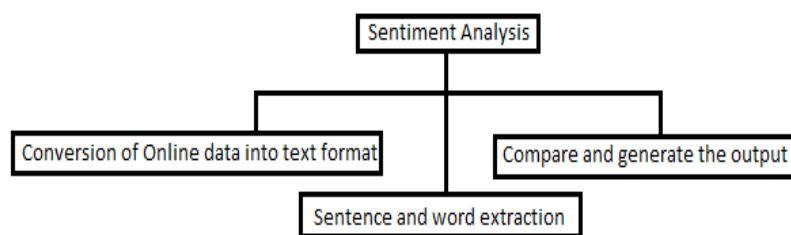
Fig3: Diagrammatic representation of the proposed System

Sentimental Analysis is a standalone Java application that provides user with options to analyze the feeling of a person in a message or phrase. Unlike other techniques , our proposed methodology and software will provide the user to analyze a document by providing the phrase or messages in it.

2.1 Design and Implementation Constraints

The proposed methodology requires a tool named facepager which is used for fetching the data from facebook. Facepager takes a bunch of Facebook page names, and then enables the user to gather all posts or likes as well as a number of other types of data; these can be exported as data files for further processing and analysis.The overall flow chart of proposed system is:-
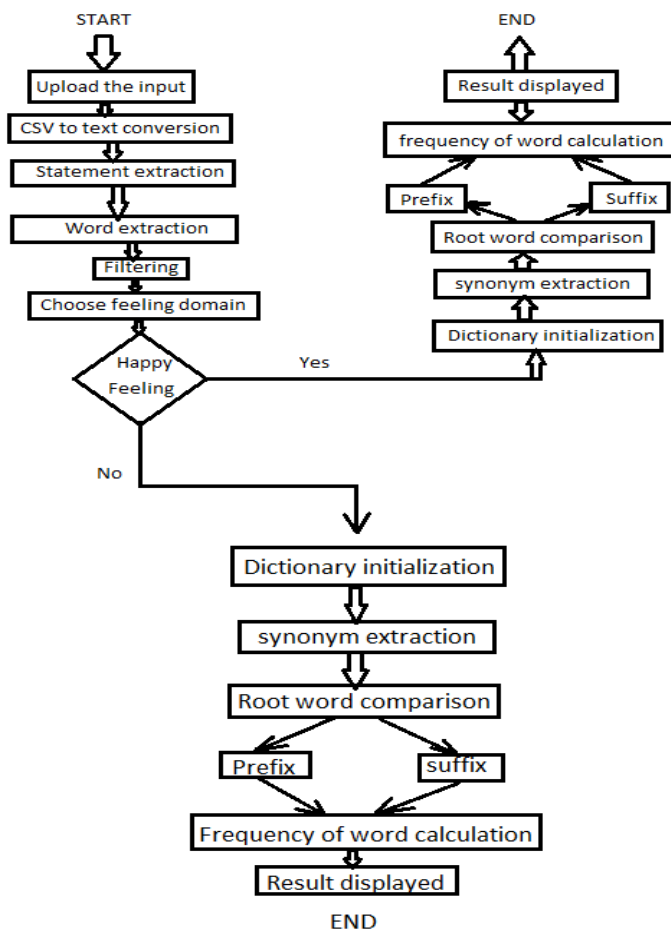
Fig4: Architecture of proposed system

3. Conclusion

This paper suggests a methodology with options to analyze the feeling of a person in different domains. It will provide the user to analyze a document by providing the phrase or messages in it. It can be used by a psychiatrist in the treatment of a patient by seeing his insight.

Sentiment analysis can be a useful tool for an organization to understand what its employees are feeling, and take remedial measures to salvage the development areas, while leveraging the strengths. Natural Language Processing can be used for processing the data with 70% to 80% efficiency.Various tools, apart from this algorithm, that helped in achieving this efficiency are wordnet (used for fetching synonyms of the words) and facepager (used for fetching data from facebook). New modules can be added without affecting the existing modules.

4. References

[1] B Pang, L. Lee, "Opinion mining and sentiment analysis", FoundTrends Inform Retriev, pp. 1-135, 2008.

[2] Medhat Walaa, Ahmed Hassan, Hoda Korashy, "Sentiment analysis algorithms and applications: A survey", Ain Shams Engineering Journal 5.4, pp. 1093-1113, 2014.

[3] Arora Piyush, "Sentiment Analysis for Hindi Language", Diss. International Institute of Information Technology Hyderabad, 2013.

[4] Kaur Amandeep, Vishal Gupta, "A survey on sentiment analysis and opinion mining techniques", Journal of Emerging Technologies in Web Intelligence, vol. 5, no. 4, pp. 367-371, 2013.

[5] Esuli Andrea, Fabrizio Sebastiani, "Sentiwordnet: A publicly available lexical resource for opinion mining", Proceedings of LREC, vol. 6, 2006.

[6] R. Hemnaath, B.W. Low, "Sentiment Analysis Using Maximum Entropy and Support Vector Machine", Semantic Technology and Knowledge Engineering, 2010.

[7] Chin Chen Chien, Tseng You-De, "Quality evaluation of product reviews using an information quality framework", Decis Support Syst, vol. 50, pp. 755-68, 2011.

[8] L Zhou, B Li, W Gao, Z Wei, K. Wong, "Unsupervised discovery of discourse relations for eliminating intra-sentence polarity ambiguities", conference on Empirical Methods in Natural Language Processing(EMNLP'11), 2011.

[9] B Heerschop, F Goossen, A Hogenboom, F Frasincar, U Kaymak, F. de Jong, "Polarity Analysis of Texts using Discourse Structure", ACM Conference on Information and Knowledge Management (CIKM'11), 2011.

[10] Hu Nan, Bose Indranil, Koh Noi Sian, Liu Ling, "Manipulation of online reviews: an analysis of ratings readability and sentiments", Decis Support Syst, vol. 52, pp. 674-84, 2012.

[11] Durie Adnan, Fei. Song, "Feature selection for sentiment analysis based on content and syntax models", Decis Support Syst, vol. 53, pp. 704-11, 2012.

[12] Kang Hanhoon, Yoo Seong Joon, Dongil. Han, "Senti-lexicon and improved Naive Bayes algorithms for sentiment analysis of restaurant reviews", Expert Syst Appl, vol. 39, pp. 6000-10, 2012.

[13] A Moreo, M Romero, JL Castro, JM. Zurita, "Lexicon-based comments-oriented news sentiment analyzer system", Expert Syst Appl, vol. 39, pp. 9166-80, 2012.

[14] Keshtkar Fazel, Diana. Inkpen, "A bootstraping method for extracting paraphrases of emotion expressions from tekts", Comput Intell, vol. 0, 2012.

 [15] Balahur Alexandra, M Hermida Jesu's, Montoyo Andre's, "Detecting implicit expressions of emotion in text: a comparative analysi", Decis Support Syst, vol. 53, pp. 742-53, 2012.

[16] Moraes Rodrigo, Valiati Joa⌣o Francisco, Wilson P. Gavia⌣o Neto, "Document-level sentiment classification: an empirical comparison between SVM and ANN", Expert Syst Appl, vol. 40, pp. 621-33, 2013.