

International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IMPACT FACTOR: 6.017

IJCSMC, Vol. 6, Issue. 4, April 2017, pg.212 – 215

Social Data Analysis

Kalyani Deore, Apeksha Dhoka, Snehal Khatod, Mansi Pagar

K.K.Wagh Institute of Engineering Education & Research, Savitribai Phule Pune University, Nashik
kalyanid1508@gmail.com, apekshadhoka61095@gmail.com, snehakhatod95@gmail.com, mansi.pagar9@gmail.com

Abstract— *The world has changed drastically in the last 10 years. An individual's opinions are no longer shared only with his or her immediate family and friends, but instead are capable of influencing the decisions of thousands or even millions of people the individual has never even met. The Internet has given the individual a platform to broadcast grievances and recommendations that can reach across the world. The attention to Web content allows the company to respond quickly to customer opinion. The sheer volume of references related to your company's products or services makes automating this task essential. Sources such as blogs, product reviews, forums and news articles can all be monitored, scored for relevance against your topics of interest, and then classified according to sentiment. Current analytical approaches in computational social science can be characterized by four dominant paradigms: text analysis (information extraction and classification), social network analysis (graph theory), social complexity analysis (complex systems science), and social simulations (cellular automata and agent-based modelling). The basic question asked in Social set Analysis is whether a given piece of text contains any subjective content (opinions, emotions, etc.) or not. This task aims to tackle this problem of differentiating between subjective and objective content. Verifying discrete polarities: Once the subjective part is determined, the next step is to determine if the content is positive or negative. This problem can be looked upon as a classification problem. Identifying subjective portions of text: The same word can be treated as subjective in one context, while it might be objective in other things. This makes it difficult to identify the subjective (sentiment-bearing) part of text.*

Keywords— *Big Social Data, Social data analysis, Text Tokenization, Text Extraction, Gate Processor, Jape Rules, Annie Plugin.*

INTRODUCTION

The idea of the project is to design and develop a system that analyses thousands of user reviews on different entities and provides a generalized opinion for that entity using natural language processing.

Social media are fundamentally scalable communications technologies that turn Internet based communications into an interactive dialogue platform. On the demand side, users and consumers are increasingly turning to various types of social media to search for information and to make decisions regarding products, politicians, and public services. As such, not many organizations are generating competitive advantages by extracting meaningful facts, actionable insights and valuable outcomes from Big Social Data analytics. It introduces a research program situated in the domains of Data Science and Computational Social Science with practical applications to Social Media Analytics in organizations. It addresses some of the important theoretical and methodological limitations in the emerging paradigm of Big Data Analytics of social media data. Finding opinion sources and monitoring them on the Web can still be a difficult task because there are a large number of diverse sources, and each source may also have a huge volume of opinionated text (text with opinions or sentiments). Selecting an attributes for sentiment classification using feature relation networks. In many cases,

opinions are hidden in long forum posts and blogs. It is difficult for a human reader to find relevant sources, extract related sentences with opinions, read them, summarize them, and organize them into usable forms. Thus, automated summarization systems are needed. Using this summarization we can recognize the importance, quality, popularity of product and services. In this system we make summarization for movie. But, we can use this system anywhere, where text analysis is required. Sentiment analysis, also known as opinion mining, grows out of this need. It is a challenging natural language processing or text mining problem.

The objective of this document is to present a detailed description of the product rating and Review Summarization. It will explain the purpose and features of the system, the interfaces of the system, what the system will do, the constraints under which it must operate and how the system will react to external stimuli.

An important application of text analytics is to automatically characterize the sentiment of documents in a variety of domains, whether it is positive, negative or neither. In this project we explore the benefits of combining domain-specific linguistic rules with data mining methods to improve both the effectiveness of your models and the efficiency of the model builder.

The Graphical representation of the generalized feedback is generated by analysing thousands of users' comment using Data mining approach and Natural language processing approach.

METHODOLOGY

1. Text Tokenization:

Text Tokenization is the process of breaking a stream of text up into words, phrases, symbols, or other meaningful elements called tokens. The list of tokens becomes input for further processing. In this case it is used to break the sentence with the help of POS tagging.

2. Text Extraction:

Text Extraction is the task of automatically extracting the relevant words or texts from a sentence. In this case it is used to extract firstly positive and negative words and the to extract any intensifiers present such as too, so, very etc.

3. Jape Rules:

JAPE is the Java Annotation Patterns Engine, a component of the open-source General Architecture for Text Engineering (GATE) platform. JAPE is a finite state transducer that operates over annotations based on regular expressions. Thus it is useful for pattern-matching, semantic extraction, and many other operations over syntactic trees such as those produced by natural language parsers.

4. Annie Plugin:

GATE includes an information extraction system called ANNIE (A Nearly-New Information Extraction System) which is a set of modules comprising a tokenizer, a gazetteer, a sentence splitter, a part of speech tagger, a named entities transducer and a co-reference tagger. ANNIE can be used as-is to provide basic information extraction functionality, or provide a starting point for more specific tasks.

ALGORITHM

1. Create JAPE rules for tagging sentiment words.
2. Initialize gate processor and Annie.
3. Execute documents in corpus to perform POS tagging
 - (a) Tokenize the document
 - (b) Perform POS tagging on every token of the document
 - (c) Also tag words with sentiments using JAPE rules (mentioned earlier)
4. Assign scores to the every sentiment based on its polarity (positive/negative)
5. Use intensifiers to gain more accurate sentiment calculation (Increment/decrement the scores further).
6. Use negation trigger to inverse the polarity of the sentiment (positive become negative and vice versa).

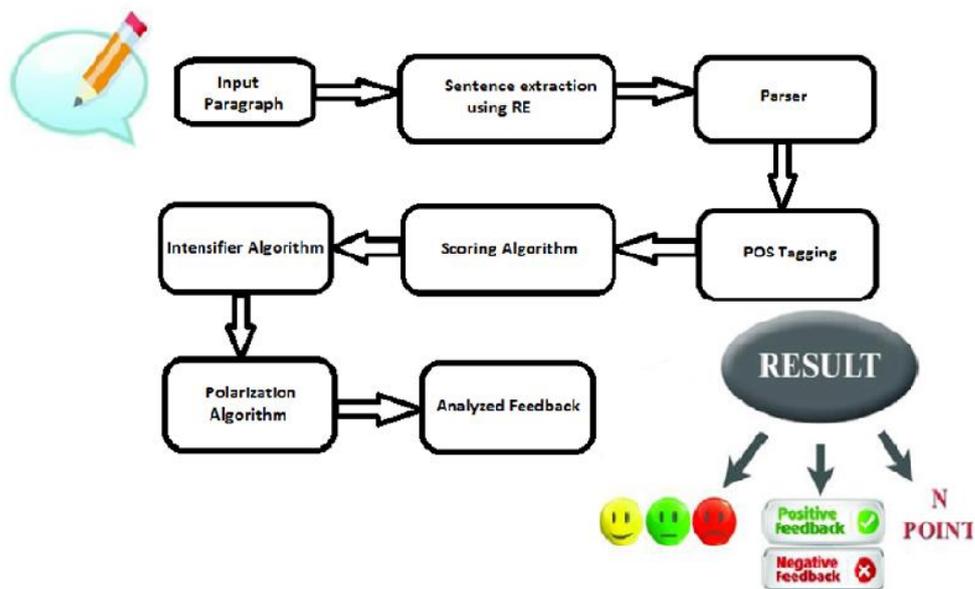


Fig. System Architecture

Description:

In order to achieve our target, we have to take following things into consideration:-

1. The graphical representation of thousands of users should be displayed properly.
2. Dictionary database should contain all the possible words.
3. There should be proper tokenization and POS tagging of the sentence.
4. Sentiment analysis should be done using proper analysis algorithm.
5. Sentences should be correctly classified according to the polarity (i.e. positive, negative, and neutral).
6. Result should be analysed properly.

The flow of the design can be presented as:

- Fetch comments from user.
- Use dictionary approaches to determine the product, user is speaking about.
- Create dictionaries for weak and strong sentiment related patterns.
- Apply strong negative sentiment patterns to the input in relative to the product.
- If not found try searching for weak negative patterns.
- Search for positive sentiment patterns in the comments with relative to the product.
- If positive sentiment pattern is found make sure that it does not have negative pattern preceding it. If found just flip the polarity of the sentiment to negative.

CONCLUSION

In this methodology we analyse user reviews on different products in market. It is used for improving the products of an organisation. The feedbacks are analysed and generalized opinions for the product is provided.

It uses data mining concepts (Text Analytics) for sentimental analysis. The key is to fully understand the whole range of issues and pitfalls of that entity, cleverly manage them and improve the quality.

Future Work:

In Future, several improvisations can be done such as adding detection of emoticons. The analysis can also be improved by localization. The detection of sarcasm can also be used to improve the analysis. We can also make use of the short forms, the detection of them is a major concern.

REFERENCES

- [1] Ravi Vatrapu, Raghava Rao Mukkamala, Abid Hussain and Benjamin Flesch, Social Set Analysis: A Set Theoretical Approach to Big Data Analytics, IEEE Access: Theoretical Foundations for Big Data Applications: Challenges and Opportunities, 2016.
- [2] R. Xia, F. Xu, C. Zong, Q. Li, Y. Qi and T. Li, Dual sentiment analysis: Considering two sides of one review, IEEE Trans. Knowl. Data Eng., vol. 27, no. 8, pp. 2120-2133, Aug. 2015.
- [3] Z. Hai, K. Chang, J. Kim, and C. C. Yang, Identifying features in opinion mining via intrinsic and extrinsic domain relevance, IEEE Trans. Knowl. Data Eng., vol.26, no. 3, pp. 447462, Mar. 2014.
- [4] R. Xia, T. Wang, X. Hu, S. Li, and C. Zong, Dual training and dual prediction for polarity classification, Proc. Annu. Meeting Assoc. Comput. Linguistics, 2013, pp. 521525.
- [5] C. Lin, Y. He, R. Everson, and S. Ruger, Weakly supervised joint sentiment-topic detection from text, IEEE Trans. Knowl. Data Eng., vol. 24, no. 6, pp. 11341145, Jun. 2012.
- [6] A. Abbasi, S. France, Z. Zhang, and H. Chen, Selecting attributes for sentiment classification using feature relation networks, IEEE Trans. Knowl. Data Eng., vol.23, no. 3, pp. 447462, Mar. 2011.