Data Mining and Machine Learning Application Design Review

Sentiment Analysis and Implementation

Assignment No. 2

By:

GIRISH KUMAR RAMACHANDRA

CWID: 888253630

CPSC 531-01 (13548)

Spring, 2019

Professor: Dr. Chun-l Phillip Chen

Department of Computer Science

California State University, Fullerton

March 26, 2019

Table of Contents

List of Figures

Abstract

Working with Sentiment Analysis is where we can classify and identify several feelings, opinions and classify them based on raw opinionated data, in which we identify the attitude of a person towards a view or product and to mainly check the bias or view and categorize it accordingly into negative view, positive view or not biased (Neutral). Natural language processing is implemented here. It can also be called as opinion mining or extraction of emotion. Text mining considered to be famous research part of Sentiment analysis. Human decision making becomes easy by this technology. While to execute sentiment analysis, we take help of various tasks like sentiment classification, detection of subjectivity, aspect term extraction, feature extraction and many more. Consumers can basically rely on Sentimentally analyzed data to review a product and buy it based on its reliability pattern. Hence, scientific communities and the business world are growing their interest in it day by day to extract data from public media and analyze them to meet the market expectations and match the standards. Knowing public views through public media is an efficient way for detecting happenings and events, while policy making can be improvised and fine-tuned, therefore many governments and business hubs are using various technologies and strategies to keep a check on them. Several problems that occur are: not able to be effective in different domains, accuracy being sometimes inaccurate and performance in sentiment analysis which rely on raw data, not able to comprehend complex sentences. The paper explains several approaches that deal with sentiment analysis and sentiment classification. Also, problems and several issues that hamper performing sentiment analysis.

Keywords: *Sentiment analysis, Opinion Mining, Sentiment classification, Social media mining, Features Selection, Semantics, Machine learning, Natural language processing.*

1. Introduction

Invention and advancements of web 2.0, surprisingly, a huge amount of data is generated which is manipulated every moment (Shahnawaz, and Astya, 2017). Sentiment analysis is a method of extracting opinionated selective data from the pool of raw data, where it allows everyone to know the views, reviews, attitudes, and emotions in their tone (Kaur, Mangat, and Nidhi, 2017). User's views and likes are collected and manipulated accordingly. This information further plays a major role in its analysis. Checking the attitude of the writer and his views about it is the task of sentiment analysis. Today internet is in the zenith and while social media plays a vital role in everyone's life in the form of microblogging sites is that the data is readily available for analysis (Kaur, Mangat, and Nidhi, 2017). And Social networking websites are so booming with data that the data is loaded with emotions, feeling, sentiments or opinions in the form of blogs, reviews, Vlogs, comments etc. (Shahnawaz, and Astya, 2017). Networking sites also use this information to their advantage to know what people think and want. Thus, opinion mining and sentiment analysis has become an important part of the business life (Kaur, Mangat, and Nidhi, 2017).

Sentiment analysis helps in checking if the product performs good or bad, and it helps them in adjusting to customer's expectations. For example, a website which reviews the laptops categorizes them based on performance and other aspects which gives a clear picture to the customer to buy a good product and they review them based on their experience (Pasarate, and Shedge, 2016). Blogs, discussions, tweets, etc. and with the advent of these data and success in making decisions business and individuals for that reason are forced to consider these data and work on them and as a result, IT giants and their research and interest in the field is also increasing (Baydogan, and Alatas, 2018).

Companies are always curious to know aspects like a Product review, product performance, QOS, Price satisfaction, likes, dislikes, expectations of customers. Sentiment analysis can identify emotions and evaluate opinions and comprehend the opinionated text (Ramanathan, and Meyyappan, 2019). We can analyze if the opinion was positive, negative or unbiased.

Web sites like Facebook, twitter are so popular theses days that they have named the field to be opinion mining as it mines the opinions. Classified as data mining and computation of linguistics. Where it identifies person's emotion, as how he behaves, what opinions he throws (Wagh, and Punde, 2018). As social networking sites are growing, implementation of sentiment analysis has entered into different sociological areas, like marketing, politics, business, sociology, environment, economy, ecology, etc. Most of these data available are unstructured; this makes analysis difficult. Sentiment analysis is a necessary technology to analyze people on the social media platform (Tyagi, and Sharma, 2017).

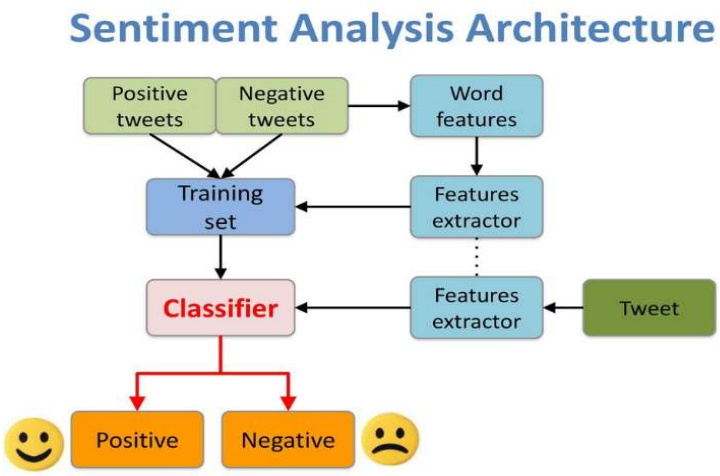

Figure 1: Sentiment Analysis Architecture. (min yuh day, 2016)

The above figure shows a glimpse of Sentiment analysis architecture from twitter data where tweets are fed into training set to sentimentally analyze them with a positive or negative result.

2.0 Literature Review

2.1 Sentiment Classification

Sentiment classification is a method of classifying text, and its objective is to do it based on the sentimental polarities of opinions, views, and emotions it contains like favorable or unfavorable, positive or negative (Kaur, Mangat, and Nidhi, 2017). This application has pulled all the interests of IT giants due to its core of selective analysis.

2.2 Machine Learning

Algorithms and statistical models when they are scientifically studied where the machine and systems correctly perform a particular task without outside instructions, depending on patterns & known facts entirely (Wikipedia). It is considered a part of artificial intelligence.

2.3 Opinion Mining

Natural language processing, mining the text, computations and biometrics is the core of opinion mining and to systematically check the affective states and other information (Wikipedia). It gives a voice and meaning to customer sayings such as reviews, responses which we get online, and other healthcare aspects for projects that cover the areas from marketing till the end user part.

2.4 Methods of Sentiment Analysis

Methods of Sentiment Analysis are categorized into four types.

Statistical approach is implementing machine learning aspects like support vector machines (SVM) algorithm, semantic orientation, latent semantic analysis, bag of words (Shahnawaz, and Astya, 2017). There are many experiments that are conducted in order to check the person who has shared his sentiment and to match with the relevant opinionated data.in several instances, grammatical relationships between the words are also used to mine the data.

Machine learning based approaches Support Vector Machine and Naïve Bayesian algorithms are major approaches used in this category. In order to do regression or classification study of data and checking the design of data is important. SVM acts as a supervised learning approach (Shahnawaz, and Astya, 2017). While Naïve Bayesian approach is based on Naïve-Bayes theorem.

Knowledge/Lexicon Based Approaches this approach based on resources about lexical and NLP, while from opinionated data, knowledge is extracted (Shahnawaz, and Astya, 2017) with this text polarity is definite.

Hybrid Approaches several researchers have experimented with this model. the method uses a ML approach and other lexical approach jointly (Shahnawaz, and Astya, 2017).

2.5 Feature Extraction Technique

The Feature extraction technique in Sentimental analysis can be split into below 4 categories.

Total Weighted Score Computing Method this model mainly considers grammar and topic correlation. Features are generally extracted, and a relation between them is found (Pasarate, and Shedge, 2016). And finally, the score is reviewed by assigning weights to the adjectives.

Neutral/Polar/Irrelevant Classification Model. Tweets are analyzed, manipulated, and their polarity is found if they are either good, bad or neutral (Pasarate, and Shedge, 2016). Application Programming Interface (API) can collect the data set processes them to filter out irrelevant ones. And the data is transformed into a feature vector.

Weighing and Aggregation Scheme

Classifying sentiment at the file level is based in scores, linguistics. Linguistic feature selection schemes are of 2 types. In the first one ‘adjectives’ and ‘adverbs’ are extracted which are preceding the selected adjective (Pasarate, and Shedge, 2016). In 2nd one, we get adverbs along with verbs and adjectives, preceding them.

Intrinsic and Extrinsic Domain Relevance Approach extracted with help of rules coordinating syntactics, features about opinions which are specific to few aspects, meanwhile not generic through IEDR are checked, then IDR and EDR are checked (Pasarate, and Shedge, 2016).

3.0 Discussion

3.1 Background and Process of Sentiment Analysis

Sentiment analysis is a method of evaluating a person's view, the emotion towards a product or a happening and it's categorized into good sentiment, negative sentiment or unbiased.

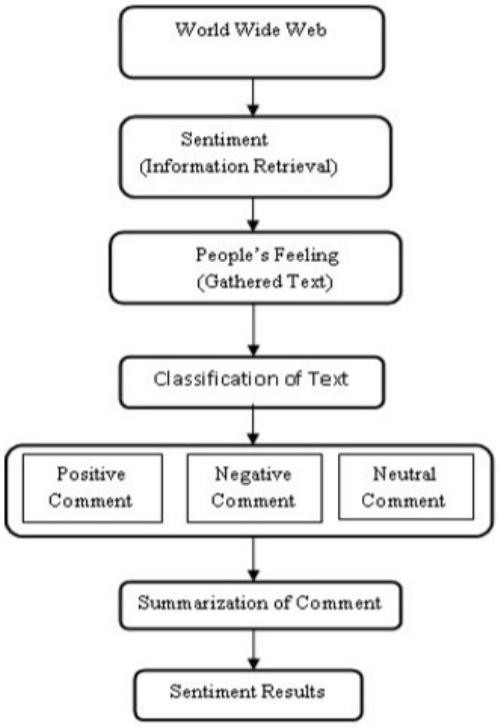

Figure 2: Process of Sentiment Analysis (Tyagi, and Sharma, 2017).

Figure above gives us a glimpse of Sentiment Analysis where input is categorized at several stages (Tyagi, and Sharma, 2017). Summarization process in the sentimental analysis has 3 main stages, retrieval of opinionated data, Sentiment classification and summarization of the Sentiment. Texts for reviews are taken from sites like Twitter, eBay, Facebook, etc. opinionated texts from many sources (Shahnawaz, and Astya, 2017), (Tyagi, and Sharma,

2017). contains subjective information. And the results are given on the basis of features selection.

3.2 Sentiment Analysis and Social Networks

The advent of internet and blogging has given voice to the words, views, and opinions that people float in the air (Tyagi, and Sharma, 2017). views such as opinions, emotions, evaluations, attitudes, etc. are mainly the feed for the study in sentiment analysis and opinion mining. The rapid growth of forums, blogging sites and microblogging sites, dating websites, matrimonial sites, and other social media tools in the crowd and with the boom in use of the web, the way things work has changed radically (Baydogan, and Alatas, 2018). Familiar social networking sites and applications like Facebook, Twitter, Instagram, YouTube, Google Play, Apple store, Vine, blog, microblog All these players highlight the "Social Media" structure.

3.3 Applications of Sentiment Analysis

Business and Organizations: Business giants need the opinions of their customers and markets; hence they spend on the resource which works on the technology (Pasarate, and Shedge, 2016).

Individuals: rely on Sentimental analysis in order to finalize the product to buy or change the opinion about happening.

Ad placement: agencies generally place more ads about the product which has good reviews.

3.4 Tools used for Sentiment Analysis

Several Sentimental analysis tools are famous that track social sphere for views, content, emotions, and sentiments. Many opensource tools are available that help in natural language processing (NLP), where they help in classifying and sourcing of data, several tools used are (Tyagi, and Sharma, 2017).

Python NLTK text processing can be done using the tool also other features are, cataloging, tokenization, stopping, tagging, parsing of data, etc .

Opinion Finder it helps in identifying individual sentences and finding polarity in them.

Apache OpenNLP based on machine learning approach, and it can process NL text in the Apache OpenNLP library.

Web Fountain which basically works on text gathering, storing indexing and querying.

WEKA is based on machine learning approach, JAVA programming language is used here, while it also has a GUI.

Ling Pipe text is linguistically processed using ling pipe, which includes extraction of text, cataloguing and cluttering of text.

Opinion Observer while user generates contents online, this tool is used to comparing and analyze the opinions they have generated.

Review Seer Tool data generated by review sites are generally manipulated by this tppl in order to segregate them based on sentiments using a Bayes classifier approach.

Red Opals based on the products; opinion orientations are reviewed using this tool.

Stanford Parser is a tool helps in post tagger and also for parsing the sentence from the NLP group.

3.5 Naïve Bayes Text Classification

Just like a probabilistic strategy, Bayesian arrangement is also managed which is Naive Bayes content classification. By making use of samples which are suitable and which imply good, bad or unbiased sentiments, also it should support the same (Kumar, Desai, and Majumdar, 2016). A dynamical model is combined with a statistically based model. And the Bayes classifier mainly works on single expressions also concentrating on pairs which match them. It classifies if the input may be right or wrong. While impartial classes will be marked +1, and -1 and 0 respectively (Kumar, Desai, and Majumdar, 2016). Basic dynamical first order system is driven by this model, where its position speaks to the mimicked enthusiastic position of its order.

3.6 Open Problems and Issues

1) The main issue we across in Sentimental analysis are comparative and objective sentences, classification accuracy and sarcasm and with flaws in these concepts will categorize the inaccurate data (Shahnawaz, and Astya, 2017). Due to which important opinionated data will be classified as neutral texts which is a major disadvantage.

2) Resources, libraries, and tools related to Sentimental analysis are always in the English language as most of the researchers prefer English. And converting these resources to different platform often causes errors and inaccurate results (Shahnawaz, and Astya,

2017). Which also puts forward domain adaptation problem. Hence the technology must be supported when it comes in other languages too.

3) Classifying the sentiment from inadequate labeled data is a difficult task and even if efforts are made it's a costly process. On the other hand, unlabeled opinionated data is comparatively easy to obtain and is inexpensive (Tyagi, and Sharma, 2017). Hence based on these many researchers work accordingly. As these techniques can be cost-effective and on the other hand it can give out almost the same results as the supervised models do.

4) Several resources like Sent WordNet, ANEW, LWIC, SentiStrength, Sentinel help in getting opinionated data and These lexicon resources are used in several places where as checking user's opinions about product reviews, checking the strength of sentiment, discovering significant emotional responses of user, patterns in online markets etc (Kumar, Desai, and Majumdar, 2016). However, the scope this method is limited to words scope, and hence the method can be a failure in case the words are not already defined in the lexicon.

4.0 Implication and Conclusion

A survey on sentimental analysis and classifying algorithms are highlighted in this paper. And the survey keeps research open for the part of Sentiment classification. Support Vector Machine and naïve Bayes are considered as good approaches in Sentiment Classification. Opinionated data in form of tweets from twitter is pretty famous. Datasets from websites like Kaggle, Amazon, twitter, yelp are commonly used and are familiar (Kaur, Mangat, and Nidhi, 2017). Mining the opinion behind user statement and brining meaning to it is the main objective of sentiment analysis (Tyagi, and Sharma, 2017).

IT giants use sentiment analysis in order to analyses data and customer views so that they can fine tune their products and business. Total Weighted Score Computing Method is commonly used while Neutral/Polar/Irrelevant Classification checks the polarity in the tweets. Weighing and Aggregation Scheme produces sentiment profile (Pasarate, and Shedge, 2016). This paper also addresses the existing problems and also the open problems associated with Sentiment analysis (Shahnawaz, and Astya, 2017). Sentiment analysis has focused on new words for the first time, and it describes new words in emotional format. and the experiment on the real data set gives us a clear picture of the present trend (Cai, Li, and Zeng, 2016). In a nutshell, a very important aspect is context consideration. Therefore, it is good if more resourses are spent on this field in form of research (Kaur, Mangat, and Nidhi, 2017).

References

H. Kaur, V. Mangat and Nidhi, *2017.* "A survey of sentiment analysis techniques," International *Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Palladam, 2017, pp. 921-925.

Shahnawaz and P. Astya, *2017.* "Sentiment analysis: Approaches and open issues," International *Conference on Computing, Communication and Automation (ICCCA)*, Greater Noida, 2017, pp. 154-158.

C. Cai, L. Li and D. Zeng, 2016. "New words enlightened sentiment analysis in social media," *2016 IEEE Conference on Intelligence and Security Informatics (ISI)*, Tucson, AZ, pp. 202-204.

S. Pasarate and R. Shedge, 2016. "Comparative study of feature extraction techniques used in sentiment analysis," *2016 International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH)*, Noida, pp. 182-186.

C. Baydogan and B. Alatas, 2018. "Sentiment analysis using Konstanz Information Miner in social networks," *2018 6th International Symposium on Digital Forensic and Security (ISDFS)*, Antalya, pp. 1-5.

E. Tyagi and A. K. Sharma, 2017. "An intelligent framework for sentiment analysis of text and emotions - A review," *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, Chennai, , pp. 3297-3302.

V. Ramanathan and T. Meyyappan, 2019. "Twitter Text Mining for Sentiment Analysis on People's Feedback about Oman Tourism," *2019 4th MEC International Conference on Big Data and Smart City (ICBDSC)*, Muscat, Oman, pp. 1-5.

R. Wagh and P. Punde, 2018. "Survey on Sentiment Analysis using Twitter Dataset," *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, pp. 208-211.

K. L. S. Kumar, J. Desai and J. Majumdar, 2016. "Opinion mining and sentiment analysis on online customer review," *2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, Chennai, pp. 1-4.

Min Yuh Day, 2016, “Sentimental Analysis in Social Media” Retrieved from https://slideplayer.com/slide/12541406/