Effectiveness and Transparency of Sentiment Analysis Tools for Academic Purposes

Iuliia Iarmolenko

Kyiv, Vyshniakivska str. 12a, Ukraine, 02140
yu.yarmolenko@gmail.com

Abstract. This work explains the importance of using sentiment analysis tools for proceeding research in different spheres of studying and knowledge mining. Here the definition of sentiment analysis is given, as well as different cases of its usage for texts processing. A comparison of effectiveness, ease of use and computing time for two open-source tools is included in this paper. This helps in making conclusions about the degree of tools' development and their possible future opportunities in a research process. Readers of this work are supposed to find out, which advantages and disadvantages do the tools have, while examining their main characteristics.

Keywords: Sentiment analysis, NLTK, Rsentiment, text processing.

Key Terms. Natural language processing, opinion mining, sentiment analysis.

1 Introduction

Subjective opinions about products, events, public speeches and many other things, that can be evaluated and generally classified as "good" or "bad" are a big interest of people today. These opinions have always been very valuable, but not so easy to collect previously. Usually those were expensive surveys conducted in order to know the public opinion on a certain issue. However, time has passed, and due to new approaches of machine learning and artificial intelligence people are able now to collect and analyze text data automatically. Public opinion is widely used in marketing, political sciences, sociological research, and other types of investigations that need analysis of people's opinions. In this paper we are going to take a closer look on a possibility to use sentiment analysis in student research and on free tools for quality analysis.

Let us first consider the definition of sentiment analysis and evolution of sentiment analysis methods. Research on sentiments is not a new thing. It was found by authors of (Mäntylä et al. 2016) that first academic studies measuring public opinion were conducted during the years of the Second World War. Those were studies motivated by political purposes. Recognizing opinions in texts, namely sentiment analysis, became automated only in the 21st century. Author of the work (Liu 2012) states, that probably the first usage of term "sentiment analysis" can be found in (Nasukawa and Yi 2003).

A quite similar, but not with exactly the same meaning term “opinion mining” most probably appeared for the first time in a work (Kushal et al. 2003).

So what is sentiment analysis itself? According to B. Liu, sentiment analysis is a series of methods, techniques, and tools about detecting and extracting subjective information, such as opinion and attitudes, from language (Liu 2010). The goal of this analysis is to identify whether the information of a text message contains positive, negative, or neutral opinion about an object or an event. The difference between sentiment analysis and opinion mining is that sentiment analysis tries to recognize a so-called “polarity” in text, or emotion, in other words. At the same time, opinion mining is a technique, which helps to divide information, which contains facts and opinions. Nevertheless, these terms are often used as synonyms.

A number of recent studies have been already dedicated towards reviewing existing sentiment analysis tools for different purposes, for example studying social media data (Abbasi et al. 2014). What differs this work is its aim to evaluate developed tools from the academic point of view. This needs assuming several points: the users are not necessarily students of technical specialties, the tools are open-sourced and scalable for large datasets, and, finally, the tools are ready to use, need no further development and can be run on the most popular operating systems. This paper is written to highlight the aspects of using sentiment analysis in academic process. Thus, the first section of main part describes how sentiment analysis techniques can be applied in studying process. The second section gives a comprehensive review of a few free sentiment analysis tools and characteristics of these tools. To reach the goals a set of following tasks need to be completed:

1. Describe the areas where sentiment analysis is used effectively in noncommercial purposes (studying, research);
2. Analyze the dynamics of sentiment analysis popularity;
3. Define free sentiment analysis tools, which can be used in academic or any other purposes;
4. Compare the effectiveness of these tools.

2 Interest Towards Sentiment Analysis in Academic Environment

People’s interest towards new computer tools in field of natural language processing is growing every day. Thus, the development of machine learning tools, cheaper and powerful computers let us analyze huge datasets. This is extremely valuable with unstructured text data, as previously people just had no similar tools to process it. The popularity of NLP in search engines grew significantly within the last 5 years. On the figure 1 a trend of search query “sentiment analysis” during the last 8 years is displayed. Another trend for a search query “opinion mining” is shown on the figure 2. The difference in trends proves that the term “opinion mining”, however, has wider meaning about collecting any kind of opinion, while sentiment analysis is a technical tool. It proves the growing interest of people towards NLP and sentiment analysis in particular.

Let us consider the most widespread noncommercial areas for using sentiment analysis, namely academic purposes. In this section we are going to make a review of successful sentiment analysis techniques used as research instruments for studying projects. Here some works in political studies, journalism, and economics are included.

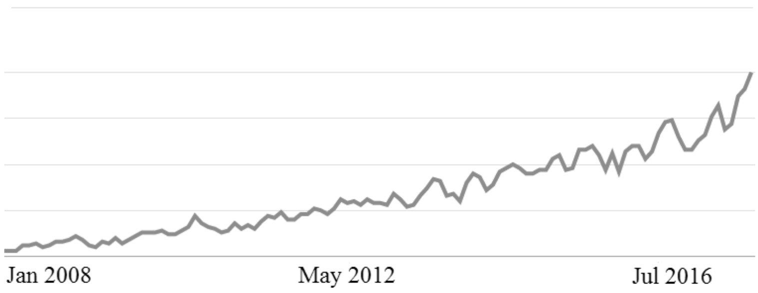

Fig. 1. The popularity of "sentiment analysis" search query in Google Trends since the year 2008[1]

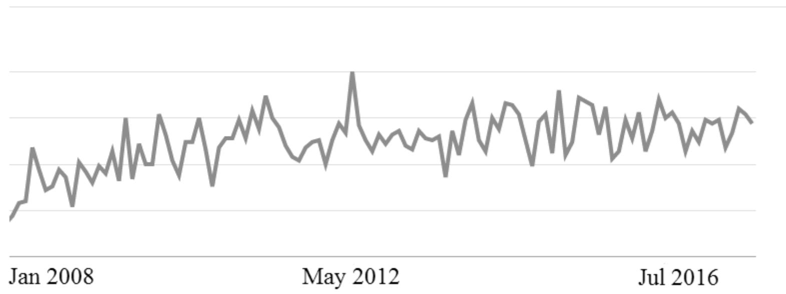

Fig. 2. The popularity of "opinion mining" search query in Google Trends since the year 2008[2]

2.1 Political Studies

In 2010 researchers from the Technical University of Munich presented the results of their work on analyzing the connection between political sentiment in tweets and the political position of candidates during the German federal election. The task raised before researchers was to find out whether micro blogging messages can be used for further political research in this area and whether "the content of Twitter messages plausibly reflects the offline political landscape" (Tumasjan et al. 2010).

To explore the issue researchers collected documents from press and election programs and compared information to stream messages from web. Due to the usage of

---

1 https://www.google.com/trends/

2 https://www.google.com/trends/

sentiment analysis it became possible to evaluate the amount of positive and negative rhetoric in candidates' speeches during debates during last 18 years. It was found out, that the most fluent authorities use mostly positive rhetoric, except for one candidate – Hort Seehofer – who is known for angry speeches and statements (Tumasjan et al. 2010).

The final results of the research are quite interesting and worth paying attention to them and the techniques used for this work. The researchers managed to prove that Twitter is a platform for political deliberation. "The mere number of tweets reflects voter preferences and comes close to traditional election polls, while the sentiment of Twitter messages closely corresponds to political programs, candidate profiles, and evidence from the media coverage of the campaign trail" (Tumasjan et al. 2010). It was also found out that sentiment profiles of separate candidates and parties overall affect the election campaigns' flow a lot. Thus we can see, how sentiment analysis is useful in political research, allowing to analyze big amounts of text data automatically.

2.2 Journalistic Studies

Journalists often work with large datasets that is why the usage of computer calculations and machine analysis is an integral part of their work. With the development of computers it became possible to use natural language processing tools in any research, as respective software became easily accessible or even free. And in terms of sentiment analysis, this technique is not only applicable to big social data, messages, tweets, etc., but also to plain texts of different genres.

A researcher from Columbia Journalism School in the work (Stray 2016) explains, why NLP tools became so important for journalists and what are the affects of this trend for journalism overall. For example, sentiment analysis tools are very effective to define propaganda in texts, manipulations with facts and so on. As an instance of successful and really discovery usage of sentiment analysis is a case with Washington Post journalists, who found out the manipulation of USAID's Inspector General office with reports (Stray 2016). The journalists proved, using sentiment analysis, that critical references were removed from initial drafts before publication.

2.3 Economic Studies

Of course, the first thing about sentiment analysis in economics that comes to mind is its usage in marketing and PR. Marketers usually use social media data to analyze the demand for a specific product, customers' reviews and so on. However, these are mostly commercial purposes that have a little in common with studying process and research. It is worth noting that the process of digitalization affected the modern economies a lot. Nowadays this impact became significant due to a strong connection between social media, other information sources and national economies. As we know, information became the most valuable resource today, and current trend shows that the strength of this connection is growing steadily.

An example of a good student research work is a degree project (Alsing and Bahceci 2015) in computer science, which investigates the connection between sentiments in tweets and fluctuations on stock market. The initial purpose of the work was to define, whether social media data can be used for predicting stock prices of a certain company. It was concluded that a set of factors, including social media activity, could be used for predicting stock markets. However, the researchers state, that sentiment analysis results can't be used as a single factor for market predicting. "This approach to stock market prediction serves better as an extra layer of complexity" (Alsing and Bahceci 2015).

In fact, there is a wide range of ways to apply sentiment analysis towards text datasets for economic type of research. Other then those mentioned approaches used in marketing and trading, sentiment analysis can be used for prices analysis (demand and supply evaluation), indexes calculation (macroeconomic indexes, based on citizens' opinions), analyzing financial news, their polarity and so on. A proper usage of NLP tools can support interesting research with statistically significant results.

3 Free Sentiment Analysis Tools

Currently Internet has a big amount of software for solving different natural language processing tasks, but the aim of this paper is to highlight those software tools, which are in free use for studying or any other purposes. It is reasonable to make a review of those programs that can work on all of the most popular operating systems (Windows, OSX, Linux), so that students and researchers with any machines can use them. This criteria also prevents us from checking low-quality software, as the most wide-spread and popular tools are usually developed for several platforms. Also a strong demand for a sentiment analysis tool is to be scalable, meaning their ability to work automatically in large datasets without human intervention.

The initial list of software found in web was reduced to just two solutions: NLP libraries for Python and R. Such programs as SentiStrength, GATE, RapidMiner, Lingpipe and others have restrictions of usage on big datasets, trial periods, time limitations for free versions, namely all those kind of restrictions that make them not scalable for using on big datasets. The advantage of open source tools is that they can be used for both academic and commercial purposes; the software is free and has a lot of contributors as well as a big users community. In this section we are going to test the most important characteristics of two sentiment analysis tools: library NLTK for Python and Rsentiment for R. Such tools' parameters as classification effectiveness, computing time, and number of classes in output is included. It is worth noting, that Stanford's CoreNLP is also a well-designed open-source tool for solving several NLP problems and sentiment analysis in particular. However, it was developed on Java and due to a lower popularity of this programming language among students of non-technical specialties it was not highlighted in the research.

NLTK is a free, open source, available for most popular operating systems. It is a community driven project, the library is being updated regularly and filled with new tools. "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"[3]. NLTK is widely used by people of different professions: linguists, engineers, researchers and students are not the whole list. The class of methods, used in our research, is called "SentimentIntensityAnalyzer" from "nltk.sentiment.vader". The output of "SentimentIntensityAnalyzer" is a probability of 4 types of sentiment in a piece of text: positive, neutral, negative, and compound. These are the types of emotions set by default for being recognized. The possibility to classify a wider range of emotions needs building and training an NLP model, which is a much more advanced problem.

On the other hand, we have a powerful and easy in usage instrument provided by R: the library Rsentiment. It is a comparably new library, which became available in CRAN repository only in May 2016 (the first release version – 1.0.4). Here is the description provided by official distributor: "Analyses sentiment of a sentence in English and assigns score to it. It can classify sentences to the following categories of sentiments: Positive, Negative, very Positive, very negative, Neutral or Sarcasm. For a vector of sentences, it counts the number of sentences in each category of sentiment. In calculating the score, negation and various degrees of adjectives are taken into consideration. It deals only with English sentences" [4]. The library has only three functions, however they give comprehensive information about a piece of text: "calculate_score" gives a numeric parameter for an emotion; "calculate_sentiment" returns an emotion found in text; "calculate_total_presence_sentiment" returns a matrix of respected counted sentiments in pieces of the whole text.

A test corpus of texts for testing these tools was taken from Sanders Analytics web source[5]. The archive contains one test file with 498 hand classified tweets and a 1.6 million classified corpora. We are going to check effectiveness on a smaller set and compare the process time on a big set. The data was processed by 1.6 GHz Intel Core i5 processor with 4 Gb of RAM.

A smaller sample of 1.6 million corpora was taken to analyze the working time of both algorithms. It took 3.3 seconds for NLTK to process 10 thousand of sentences. To compare, Rsentiment processed the same 10 thousand during 41 minutes! This is a huge difference in methods, and disadvantage of Rsentiment's computing time is significant, especially when working with massive datasets.

Now let us consider the accuracy of classification with main metrics: recall, and precision. Precision is the ratio of truly classified documents to selected elements, while recall is the number of truly classified to relevant elements. The formulas are the following:

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

[3] http://www.nltk.org/
[4] https://cran.r-project.org/web/packages/RSentiment/RSentiment.pdf
[5] http://www.sananalytics.com/lab/twitter-sentiment/

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

where TP – true positive, FP – false positive, FN – false negative.

The results of comparison are organized in the table 1.

Table 1. Comparison of sentiment analysis tools

	NLTK	Rsentiment
Precision	0.56	0.31
Recall	0.49	0.29
Computing time on 10,000	3.3 seconds	41 minutes
Variety of classes	Positive Negative Neutral Compound	Positive Negative Neutral Sarcasm

The comparison has showed that NLTK is a much better tool for sentiment analysis than Rsentiment. The problem of both classifiers is that it has 4 classes to recognize, while the test dataset was represented by two classes: positive and negative (the neutral class was removed in order to calculate precision and recall correctly). That means that, in fact, the classification was not useless since we have to adjust our results for the fact, that there were 4 classes in outcome.

4 Results and Discussions

NLTK and Rsentiment are two libraries including sentiment analysis methods for classifying sentiments in text data. Both of them are able to recognize four classes of emotions. However, computing time for these methods differs a lot, and that might be a problem for Rsentiment package, which takes about half a second to process one short piece of text (a tweet is a message of length no more than 140 characters). Moreover, the results of classification showed, that NLTK has really high results of classification, considering that it gives four classes and not two as an outcome. However, it is worth noting that Rsentiment is a comparably new tool, and later versions of this package are expected to be more effective.

The current research has several limitations. First of all, only fully developed methods were tested as sentiment analysis tools. Normally a deep learning for identifying sentiments in texts needs a machine learning approach. Those tested tools are universal for texts and they were learned on certain datasets that might not consider such aspects as sphere and terminology for a specific field of studying. Secondly, we used tweets to check the effectiveness of our tools, which have a lot of slang, mistakes, and other kind of bias that might affect the results significantly.

5 Conclusions

Our research showed that sentiment analysis techniques are really important in studying process for specialists from many spheres. It is already successfully used in political science and journalism, economics and marketing, etc. We can find a lot of examples of effective implementation of sentiment analysis tools, which helped researchers to recognize emotions in big unstructured text sets. Considering this, the author tried to take a closer look at the most popular sentiment analysis tools: NLTK library for Python and Rsentiment for R. It was found that NLTK is a more precise and much faster tool for sentiment analysis, however the results of classification are not perfect. The very same open-source languages Python and R can be used for creating better classifiers on text corpuses specific for respective areas of studying.

References

1. Abbasi, A., Hassan, A., Dhar, M.: Benchmarking Twitter Sentiment Analysis Tools. In: Calzolari N., Choukri K. (eds) LREC ′14: Ninth International Conference on Language Resources and Evaluation, Reykjavik, May 26-31 (2014).
2. Mäntylä, M.V., Graziotin, D., Kuutila M.: The Evolution of Sentiment Analysis - A Review of Research Topics, Venues, and Top Cited Papers. (arXiv:1612.01556v1 [cs.CL]) (2016)
3. Liu B.: Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers, San Rafael, CA (2012)
4. Nasukawa T., Yi J.: Sentiment analysis: capturing favorability using natural language processing. In: Gennari J, Porter B. (eds) K-CAP ′03: 2^{nd} International conference on Knowledge capture, Sanibel Islands, FL, October 23-25 (2003)
5. Kushal D., Lawrence S., Pennock D.M.: Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In: Hencsey G, White B (eds) WWW ′03: 12^{th} International conference on World Wide Web, Budapest, May 20-24 (2003)
6. Liu B.: Sentiment Analysis and Subjectivity. In: Indurkhya N, Damerau FJ (eds) Handbook of Natural Language Processing, 2^{nd} edn. Chapman & Hall, Boca Raton, FL, p. 627-661 (2010)
7. Tumasjan A., Sprenger T.O., Sander P.G. et al.: Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. In: Hearst M. (ed) AAAI ′10: 4^{th} International conference on Weblogs and Social Media, Washington, D.C., May 23-26 (2010)
8. Stray J.: What do Journalists do with Documents? Field Notes for Language Processing Research. Stanford Journalism Department, https://journalism.stanford.edu/cj2016/files/What%20do%20journalists%20do%20with%20documents.pdf (2016)
9. Alsing, O., Bahceci O.: Stock Market Prediction Using Social Media Analysis: Degree Project in Computer Science. Stockholm, http://www.diva-portal.se/smash/get/diva2:811087/FULLTEXT01.pdf (2015)
10. NLTK 3.0 Documentation, http://www.nltk.org/
11. Package “Rsentiment” Documentation, https://cran.r-project.org/web/packages/RSentiment/RSentiment.pdf
12. Sanders Analytics web source, http://www.sananalytics.com/lab/twitter-sentiment/