IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

Comparison of Web Services for Sentiment Analysis in Social Networking Sites

To cite this article: Ain Balqis Md Nor Basmmi *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **884** 012063

IOP Conf. Series: Materials Science and Engineering **884** (2020) 012063 doi:10.1088/1757-899X/884/1/012063

Comparison of Web Services for Sentiment Analysis in Social Networking Sites

Ain Balqis Md Nor Basmmi[1], Shahliza Abd Halim[1] and Nor Azizah Saadon[1]
[1] School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

E-mail: ainbalqis96@gmail.com

Abstract. With various type of web services available, it is hard to identify and compare which of the free access web services work best in analysing sentiment of extremist content in social networking sites. For that purpose, a generic approach by working with API of web service using PHP programming language is used to test each dataset that was extracted based on the keyword 'extremism'. Data from both Twitter and Facebook has been used as these two are the most powerful platforms for expressing one's feeling. The comparison for web service is done based on the analysis of its accuracy, precision, recall and f-measures in obtaining the lowest score of mean square error (MSE). Four sentiment analysis web services are used which are Sentiment Analyzer, Aylien, ParallelDots, and MonkeyLearn. From the comparison, MonkeyLearn obtained the best final results among all web services with the lowest MSE score of 14%. For the benefit of other researchers, the finding of this will reveal the suitable web service for analysing sentiment issues as critical as extremism.

1. Introduction

In our current connected society, the amount of data online has increased abruptly every year. Social networking sites are commonly used by people to express their feelings either positively or negatively about certain issues. These sites play an important role as platforms for user to create and share point of views [1]. Nowadays, terrorism is considered as a part of trend in extremist violence. With the increasing rate of users expressing their own view related to extremist on social networking sites, methods to counter extremism need to be strengthen. At this rate, the likelihood for the researchers to be able to efficiently gather representative data from web manually is impossible. Moreover, the amount of data gathered are usually too vast to be analysed. Therefore, an automated and more simple process to gather and analyse data is necessary.

Sentiment Analysis (SA), also known as Opinion Mining or Emotional Artificial Intelligence, generally aims to analyse people's attitude towards element such as topic or emotional reaction. Serrano et al. [2] stated that SA is a concept that includes many tasks, including sentiments extraction, sentiments classification, subjectivity classification, opinion summarization or opinion spam detection. It is also used to detect sentiment polarity, which falls under positive or negative opinion [3], and to identify the expression of emotion.

While SA is a concept of analysing people's emotion, web service is used as a tool to determine SA. Using a list of web services, this study will focus on finding the best web service based on its capabilities in classifying sentiment polarity based on its analysis score which falls under positive, negative, or neutral. In other words, this research aims to find a web service that functioned best in giving the most accurate result during analysing sentiment. For that purpose, each web service will be differentiated in term of its functionalities and techniques used to generate results. Besides, the results obtained during analysing sentiment will also be taken into account as the best web service will generate the least mean square error (MSE). Therefore, the focus of this paper is to propose an explicit process for extraction and comparing the web service for sentiment analysis. To be specific, this involves building a PHP page with authentication of Application Programming Interface (API) of each web services. Furthermore, the comparison will be based on quality attributes such as accuracy, precision, recall and MSE score of the web services in analysing sentiment.

2. Related Works

Serrano-Guerrero et al. [2] have done a study on SA to analyse people's reaction on movies, Twitter and Amazon's product. In their paper, it has been stated that SA is a concept that includes many tasks, including sentiment extraction, sentiment classification, subjectivity classification, opinion summarization or opinion spam detection. A refined categorization of well-known techniques of SA is used in conjunction with the task as referred to in the work done by Medhat et al. [4]. They highlighted the use of lexicon-based and machine learning approach.

Besides, Kharde et al. [5] have done a survey and comparative study which focuses on SA of Twitter data and techniques of opinion mining that includes cross-domain and cross-lingual methods and certain evaluation matrices. The study provides a general model for SA based on the analysis of Twitter data. This survey also focuses on machine learning and lexicon-based approach, and the results showed that combining machine learning method with opinion lexicon method is better in helping to improve the accuracy of sentiment classification as it consists of dictionary-based and corpus-based terms as sub-classification. The approach is predominantly based on sentiment lexicon which is a set of pre-compiled terms, phrases, and even idioms developed for genre of communication.

In term of web service, Serrano-Guerrero et al. [2] have also done a review and comparative analysis of web service using 15 different web services. In comparing the capabilities of these tools for sentiment analysis, each of them performed the most important task which is sentiment classification rating prediction using 4 types of dataset; Twitter dataset, movie dataset, Amazon whole reviews, and amazon utterances. With respect to the main tasks, the ability to classify the polarity of text can be detected. Unlike the previous research, study made by Kharde et al. [5] have highlighted four main levels of SA which are word level SA, sentence level SA, document level SA, and feature-based SA. A small relation can be seen between these two papers when some of the levels stated earlier has been proven to be used by certain Web Service in Serrano-Guerrero et al. [2] study.

Last but not least, API is used in majority of applications such as Facebook, Twitter, and web service as it enables various platform to communicate [6]. With individual web services API, we are able to analyse sentiment with the functionality offered by the specific web service. Besides API of web service that will be used during authentication in PHP page, Twitter also provides an API in the form of REST API, Search API, and Streaming API. As stated by Xing Fang et al. [7], developers can select status data and user's information with the REST API. Search API allows developers to query Twitter content specifically while Streaming API collects real-time content from Twitter. However, research made by Mahmoud Al-Ayyoub et al. [8] proved that although all APIs are provided by Twitter, there are many limitations in using them. Additionally, Streaming API is proven to be more flexible compared to REST API, making it the preferred tool for this process. However, developers can mix these APIs to have a strong foundation in analysing sentiment.

3. Extraction and Comparison Process of Sentiment Analysis

In order to extract and compare web services for sentiment analysis, we proposed four phases as shown in Figure 1.

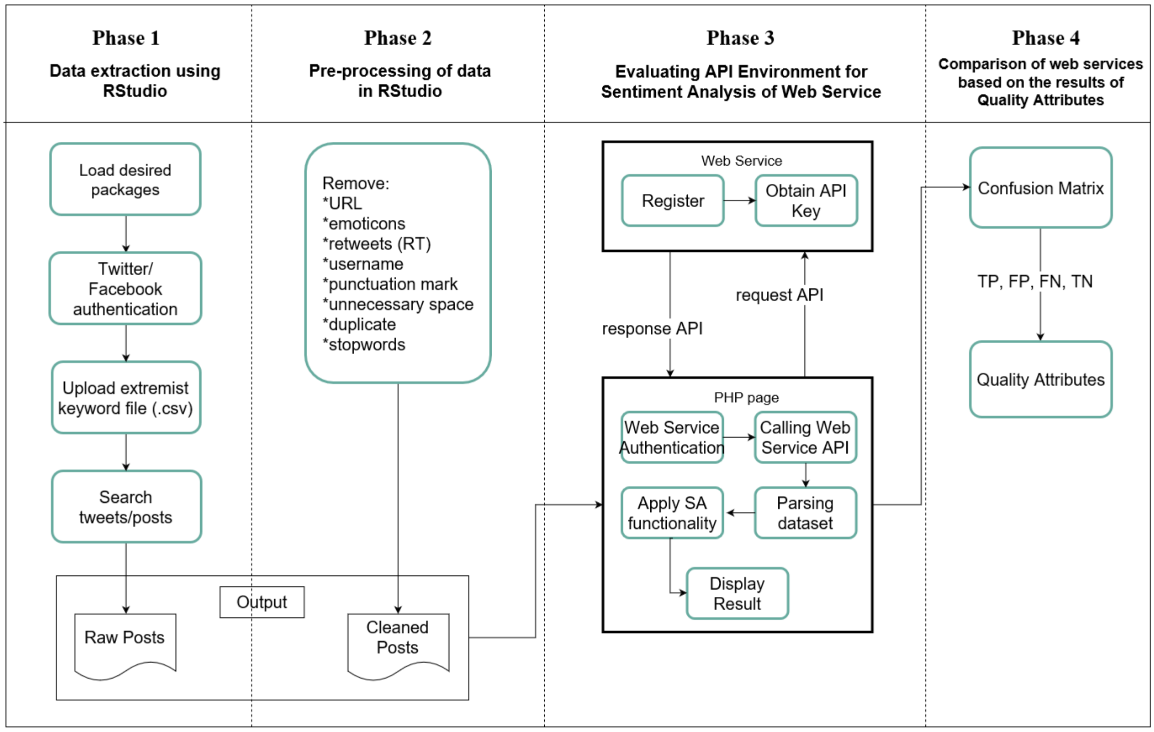

Figure 1. Sentiment Analysis Extraction and Comparison Process.

3.1. Data Extraction using RStudio

RStudio, an open source data analysis software will be used for data gathering. For Twitter dataset, there are 4 main components that are needed prior to data gathering which are: Consumer Key (API Key), Consumer Secret (API Secret), Access Token, and Access Token Secret. These components can be obtained once user has login into their Twitter account. While for Facebook dataset, the data for Access Token can also be accessed once the user has already logged in. Searching of tweets and posts involves a keyword file which contain a list of frequently searched terms related to extremism obtained from study that has been done by Ahmed et al. [9].

3.2. Pre-processing of Data

RStudio also has the ability to pre-process data. It is able to clean the data before further testing are conducted. It helps in transforming data into a form that is easier to work with. Only a few types of pre-processing activities were done depending on the condition of dataset. Among the pre-processing done are removing URL, removing re-tweet, username, punctuation marks and also unnecessary spaces.

3.3. Evaluating API Environment for SA of Web Service

The focus of this phase is to work with API of web service using PHP programming language. In this research, PHP page with API Key has been created as a platform to test all the web services. Each web service will have an API key and each key will be implemented for authentication in the PHP page. Sentiment Analyzer is the first lexicon-based SA web service to be tested that uses VADER (Valence Aware Dictionary and Sentiment Reasoner) to understand feelings in a sentence. This web service works by passing words to it and returning the word-related feeling. Each word is processed and then is labelled as a feeling by identifying the most prevalent sentiment [10]. The most crucial step in evaluating this web service is to include the package via composer. This package will later be initialized in the main source code to be used for testing. Aylien is a web service that offers a powerful and flexible AI-driven content analysis solutions that deliver NLP's power to the masses. Aylien is also capable of handling multiple documents [11]. The accuracy of web service's performance was measured after applying SA approaches to text documents. Unlike Sentiment Analyzer, Aylien requires its user to register in order to obtain an API key for analysing sentiment. In Aylien, text to be tested are stated in the back-end source code where the polarity of the sentiment will be analysed. ParallelDots is the third web service that provides SA features with very precise analysis of overall emotion of text content. It offers wide-ranging applications for review and social media, ranging from marketing to customer service [12]. ParallelDots detect feelings, keywords or phrases, and languages from text posted by individuals. It will generate three different scores which are positive, negative, and neutral. MonkeyLearn is the fourth web service which provides machine learning platform to obtain relevant text data. The aims are to retrieve and classify text information for specific needs and to integrate it easily, quickly, and cost-effectively into their own platforms and application [13]. MonkeyLearn also allows users to create their own custom classifier to train their model using its machine learning algorithm. Implementation of MonkeyLearn requires API key and model_id. Both can be obtained once the user registers in MonkeyLearn account.

3.4. Comparison of Web Services based on the Results of Quality Attributes

Confusion Matrix is the combination of sentiment polarity obtained from the web services and based on human prediction. In other words, each data will have two different or same polarity results. For example, for a sentence, web service may show a positive polarity, but we might get a different result based on human prediction. Confusion Matrix in [14] were used to describe the performance of dataset classification which consist of True Positive (TP) that indicates positive reviews and was classified as positive by the classifier, while False Positive (FP) represents positive reviews but was not classified as positive by the classifier. Similarly, True Negative (TN) indicates negative reviews and was classified as negative by the classifier, while False Negative (FN) represents negative reviews but was not classified as negative by the classifier. Results of dataset from Confusion Matrix will then be used to calculate the quality attributes. In this research, the quality attributes that we are trying to focus on are the accuracy, precision, recall, f-measures, and mean square error (MSE) of sentiment result for each web service. Table 1 summarizes all related quality attributes.

Table 1. Quality Attributes.

Name	Description	Example
Accuracy	To measure how accurate is the classifier	$= \left(\frac{TP+TN}{TP+TN+FP+FN}\right)$
Precision	To measure how precise is the results generated by the classifier	$= \left(\frac{TP}{TP+FP}\right)$
Recall	To measure the ability to find all relevent cases within dataset	$= \left(\frac{TP}{TP+FN}\right)$
F-measures	To measure the average of recall and precision	$= \left(\frac{2 * Precision * Recall}{Precision+Recall}\right)$
Mean Square Error (MSE)	Average square of errors	$= n^{-1} \sum_{n=1}^{n} (x_i - y_i)^2$

4. Results and Discussions

Sentiment Analyzer, Aylien, ParallelDots, and MonkeyLearn are web services that are chosen for this exploratory study. There will be two types of sentiments which are human prediction and machine prediction. The same data and text prediction sentiment will be used for each web service. For sentiment based on human prediction, there are 777 negative sentiments and 223 positive sentiments altogether.

Due to different characteristics, it is not easy to compare these four tools. All texts for each document were processed and the entire document is labelled as human prediction of either negative or positive. For the purpose of calculating the final result, the texts are submitted to each web service and the score obtained by each text, considered as machine prediction, was aggregated using arithmetic mean to obtain the final score.

In general, extremist content in social networking sites has been proven to contain more negative posts compared to positive posts. All web services show a higher amount of negative sentiments than positive sentiments and MonkeyLearn works best in detecting negative sentiments. Table 2 shows the overall result of comparison for all web services.

Table 2. Overall Result.

Criteria	Sentiment Analyzer	Aylien	ParallelDots	MonkeyLearn
Total Reviews	1000	1000	1000	1000
Positive Reviews	308	444	302	298
Negative Reviews	692	556	698	702
TP	208	540	219	267
FP	100	93	97	98
FN	100	16	83	31
TN	592	351	601	604
Accuracy (%)	80	89	82	87
Precision (%)	68	85	69	73
Recall (%)	68	97	73	90
F-measures (%)	68	91	71	81
Positive MSE (%)	10	43	18	17
Negative MSE (%)	19	9	10	10
Total MSE (%)	15	26	14	14

From the table, it is clear that Aylien presents a good accuracy with 89%. It also obtains the highest result of quality attributes. On the other hand, Sentiment Analyzer shows the worst performance in term of quality attributes. Despite the fact that it generates quite a high number of negative sentiments, other evaluation measure for Sentiment Analyzer are still the lowest. It proves that the evaluation measures do not influenced by the number of reviews. On the other hand, ParallelDots and MonkeyLearn also provides a good result and achieves second best in term of quality attributes scores.

In terms of MSE, the lower the MSE, the better the web service in SA. This is because, the closer the MSE result to true zero, the better it is in weighing the intensity of sentiment [2]. It can be observed that both ParallelDots and MonkeyLearn present the same MSE score and also the lowest errors with 14%. On the other hand, the worst results were computed by Aylien with the highest MSE of 26%. In addition, considering the negative sentiments, ParallelDots and MonkeyLearn obtain the same results, which is 10%. This result indicates that these two tools are the best in weighing the intensity of negative sentiments. Therefore, looking into the overall result for each web service in Table 2, it is necessary to point out that MonkeyLearn is the best web service due to its high accuracy and other quality attributes results, as well as having the lowest result of MSE.

5. Conclusion

This research is a way of achieving awareness of extremism that has been thriving on social media nowadays. There are not many researches that has been done to analyse on this issue. As this research focusses on applying the needs to capture extremist content, it may help other researchers to find an alternative solution on analysing sentiment that are related on extremism. Researchers may have sufficient information regarding the services offered by these web service and outcomes anticipated from them. As the functionality of SA web service continues to evolve, it should be noted that there are many challenges waiting for these web services. They might suffer from problems like analysing a vast range of data at once instead of a single text; or to differentiate the web service from different aspect other than sentiment polarity such as using its geo-location information. Geo-location helps in identifying and estimating the location where the particular post originated from. From this, we can see which part of the world are hotly discussing on a particular topic. To a certain extent, these two suggestions could be alleviated if there exist a web services that are capable of analysing several data simultaneously. In conclusion, this research found that MonkeyLearn is the best web service in classifying sentiment of extremist content in social networking sites. This is due to its high-quality attributes result and lowest result of MSE.

Acknowledgements

We would like to thank the Ministry of Education (MOE) Malaysia for sponsoring the research through the Fundamental Research Grant Scheme (FRGS) with vote number 5F080 and Universiti Teknologi Malaysia for providing the facilities and supporting the research. In addition, we would like to extend our gratitude to the lab members in the EReTSEL Lab, School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia for their invaluable ideas and support throughout this study.

References

[1] Kaplan A M and Haenlein M 2010 Users of the world, unite! The challenges and opportunities of Social Media *Business horizons* **53(1)** 59-68
[2] Serrano-Guerrero J, Olivas J A, Romero F P and Herrera-Viedma E 2015 Sentiment analysis: A review and comparative analysis of web services *Information Sciences* **311** 18-38
[3] Liu B 2012 Sentiment analysis and opinion mining *Synthesis lectures on human language technologies* **5(1)** 1-167
[4] Medhat W, Hassan A and Korashy H 2014 Sentiment analysis algorithms and applications: A survey *Ain Shams engineering journal* **5(4)** 1093-1113
[5] Kharde V A and Sonawane P 2016 Sentiment Analysis of Twitter Data *International Journal of Computer Applications* **139(11)** 5-15
[6] Littman J, Chudnov D, Kerchner D, Peterson C, Tan Y, Trent R and Wrubel L 2018 API-based social media collecting as a form of web archiving *International Journal on Digital Libraries* ***19*****(1)** 21-38
[7] Fang X and Zhan J 2015 Sentiment analysis using product review data *Journal of Big Data* **2(1)** 5
[8] Al-Ayyoub M, Essa S B and Alsmadi I 2015 Lexicon-based sentiment analysis of Arabic tweets *IJSNM* **2(2)** 101-114
[9] Ahmed M and George F L 2017 *A War of Keyword* (London: Tony Blair Institute for Global Change)
[10] Garay J, Yap R and Sabellano M J 2019 An analysis on the insights of the anti-vaccine movement from social media posts using k-means clustering algorithm and VADER sentiment analyser *IOP Conference Series: Materials Science and Engineering* **482(1)** 012043

[11] Alsaqer A F and Sasi S 2017 Movie review summarization and sentiment analysis using rapidminer *2017 International Conference on Networks & Advances in Computational Technologies (NetACT)* 329-335

[12] Zhang Z 2018 Sense-Recommendation System based on Affective Interpretations of Social Media Posts: A Proposed User Interface Design *Rochester Institute of Technology*

[13] Wang Z, Bai G, Chowdhury S, Xu Q and Seow Z L 2017 TwiInsight: Discovering Topics and Sentiments from Social Media Datasets arXiv preprint arXiv:1705.08094

[14] Tripathy A, Agrawal A and Rath S K 2016 Classification of sentiment reviews using n-gram machine learning approach *Expert Systems with Applications* ***57*** 117-126