A Study of Sentiment Analysis in Context of Reviewing Product: Method and Tools

[1]Nor Anis Asma Sulaiman , [1]Nor Alina Ismail, [1]Siti Norbaya Daud

School of Information Technology, SEGi University

anissulaiman@segi.edu.my

Abstract. Sentiment Analysis (SA) examines the hassle of studying texts, like posts and reviews, uploaded by way of customers on microblogging platforms, forums, and electronic agencies, concerning the evaluations they have got about a product, service, event, man or woman or idea. The common used of SA is classifying text to a several class such as positive, neutral and negative term classification. This paper aims to compare the techniques being used to examine SA especially for reviewing product or services. These techniques are not only being applied in review product prediction, but they were applied in different filed such as production, sales and marketing. Based on the comparison of each model and their uses, lexicon-based approach is the best solution for classifying SA.

Keywords: Sentiment Analysis, E-Services, Opinion Mining, Machine Learning

1. Introduction

Sentiment is a term for sensitivity to feelings of emotions, thoughts, ways of thinking about something or idea based on feelings about a situation (Rezaei & Jalali, 2017; Park & Seo, 2018; Rizwana, Phil, Scholar, & Kalpana, 2018; Zvarevashe & Olugbara, 2018). Sentiment analysis (SA) or called opinion mining (OM) is a sub-field of machine learning that has become an

11

important area of today's research to understand people's opinions about something by analysing a large amount of information available in the web nowadays. There are millions of people express their thoughts on various services or products (Priya & Sathiaseelan, 2017).

In addition, SA are the language processing tasks that use computational approaches and stratification of a person's views provided in quantitative methods such as text or video and audio primarily pertaining to one's thinking about a situation or idea of a particular topic. These thoughts and ideas are measured using a "polarity" scale of texts that are called positive, neutral and negative (Salinca, 2016; Zhang & Zheng, 2016; Akter & Aziz, 2017; Pradesh & Pradesh, 2017).

Traditionally, if someone wants to buy a product or use a service, they will ask the friends or relatives opinion first. Now users only review from the internet and then choose the highest ranked products and services (Gupta, Tiwari, & Robert, 2016; Rana & Singh, 2016). SA can provide an overview of an individual's experience or thinking about a product or service. Active feedback from a user not only for certain parties to analyse consumer satisfaction about the product, quality, cost and competitor monitoring but also very useful for users who want to research products or services before making any purchase (Ghorbel & Jacot, 2011; Jain & Katkar, 2015; Madhoushi, Hamdan, & Zainudin, 2015; Pasarate & Shedge, 2016).

SA is widely used primarily on social or online based sites to get comments on customer opinions or customer notice of the product, especially in terms of cost and quality of the product. (Anshuman, Rao, & Kakkar, 2017). There are many challenges in the analysis of sentiment. The first is that the word of opinion which is considered positive in a situation can be considered negative also in other situations. The second challenge is that people do not express their opinions in the same way. Most reviews will have positive and negative comments, which are fairly easy to handle by analysing the verses one at a time. However, in more informal media such as twitter or blogs, people who are more likely to combine different opinions in the same sentence

to make it easier for others to understand, but it is more difficult for the computer to decipher it (Routray, Kumar Swain, & Prava Mishra, 2013; Ankitkumar, R. R. Badre, & Mayura Kinikar, 2014; Pasarate & Shedge, 2016). Figure 1 shows the basic responsibility of SA or OM for analyse and classify documents or statements with appropriate text polarity categories.
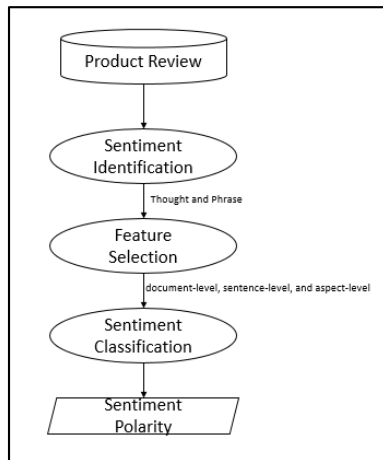

Figure 1. Sentiment Analysis Process on Product Review

Sentiment analysis is not solely applied to product reviews but can be applied to stock related news articles, political debates, individual work achievements, E-Commerce websites and also detection of spam (Gaur & Pruthi, 2017). Data set at stage sentiment identification will be processed and categorized whether the data is a category of thought and user opinion or phrase data set. Meanwhile for the classification stage of sentiment as shown in **Figure 1**, there are three main classifications in Sentiment analysis namely document-level, sentence-level, and aspect-level (Schouten & Frasincar, 2016; Gaur & Pruthi, 2017; Sarawgi, 2017) .

Document – Level

In this method the entire document is considered a single entity and the entire document is analysed at one time only. Sometimes the results provided by this approach are not appropriate. Documents that give a positive opinion of an entity do not imply that the author only has a positive comment on all

13

the features of that particular entity, as well as documents that give negative opinions about an entity do not indicate that the author is totally negative for all features of the entity. In conventional text, the author expresses positive and negative opinions about the entity and its nature.

Sentence – Level

In this method the document is broke down into sentences and then each verse is considered as a single entity and one sentence is analysed at one time. The results generated by this approach are better than document level and more subtle. The majority of today's techniques, try to establish polarity text for the entire document, paragraph, sentence or text regardless of the specified entity.

Aspect – Level

The fine-grained analysis is done in this method. The main focus is the aspects and features of the product. The result of this approach is to know the sentiments about the features of items. Most approaches when trying to determine the overall sentiment of the punishment regardless of the nature and characteristics of the target entity (For example, Mobile phones, Laptops) and their properties (size, memory, screen, and camera). In the aspect based on an analysis of text polarity sentiments for each given aspect. Aspect-Based Sentiment Analysis has a wide range of applications in different fields such as services, movie surveys, restaurants, movies and computers. It uses texts expressed in various ways such as comments, feedback, forum comments, messages and discussions.

The following section will explore the most frequently used SA techniques from past research and appropriate tools used for analysing sentiment approach.

2. Sentiment Classification Techniques

Sentiment classification techniques can be roughly divided into machine learning approach, lexicon based approach and hybrid approach (Kinge, Nandedkar, & Narkhede, 2015; Sengottuvelan & Karthik, 2017). Machine Learning Approach (ML) uses the famous ML algorithm and uses linguistic

14

features. The lexicon-based approach depends on sentiments, a collection of known and precompiled sentiment terms. It is divided into dictionary-based approaches and corpus-based approaches that use statistical or semantic methods to find polarity of sentiment. The Hybrid approach combines both approaches and it's very familiar with sentiment lexicon who plays an important role in most methods.

The classification text method using ML technique can be divided into two methods ie supervised and unsupervised learning methods. The Supervised Learning method uses a large number of labelled training documents. The supervised machine learning algorithms which deals more with classification includes Linear Classifiers, Logistic Regression, Naïve Bayes Classifier, Perceptron, Support Vector Machine; Quadratic Classifiers, K-Means Clustering, Decision Tree and others (Kotsiantis, 2007; F.Y et al., 2017; Bzdok, Krzywinski, & Altman, 2018). Meanwhile the unsupervised learning method is to identify hidden patterns in non-label input data. Data is grouped into classes and algorithms are made to differentiate. This Unsupervised Learning method also refers to the ability to learn and organize information without providing an error signal to assess potential solutions. In addition it is used when it is difficult to find documents that are in the form of raw data (Sathya & Abraham, 2013; F.Y et al., 2017;).

The lexicon-based approach depends on finding the opinion or lexicon idea used to analyse the text. A dictionary-based approach that relies on finding basic words, and then searching for their synonyms and antonyms. The corpus-based approach begins with a basic list of words and then searches for words that have the same meaning or less in the big corpus to help find a more specific meaning for the base word (Augustyniak, Szymánski, Kajdanowicz, & Tuliglowicz, 2016; Al-Hussaini & Al-Dossari, 2017). This can be done using statistical or semantic methods.

In Hybrid Techniques, both machine learning combinations and basic lexicon approaches are used. Researchers have proven that this combination provides improved performance classification. The main advantage of hybrid

15

approaches is that it uses the best lexicon / learning symbolic of the combination of both methods i.e. ML and Lexicon techniques. This makes the readability of the data set tested high in accuracy results from the applied algorithm (P & Patel, 2015; Aggarwal & Gupta, 2017) .

3. Tool Used in Sentiment Analysis

There are a variety of sentiment analysis tools available now. Some are stand-alone; some are attached to specific social media applications. The tools used in the process of tracking the opinion or polarity from the user's generated contents are:

Table 1. Tools for Analysis Sentiment

No.	Tool Name	Description
1	Apache Hadoop	Is an open source framework enables quickly gain insight from massive amounts of structured and unstructured data. Also supports powerful and scalable directed graphs of data routing and transformation.
2	BrandWatch	Get deep insights into consumer opinion on any topic from across the social web. Also can deliver a huge amount of data for analysis. Its extensive filters and customizable reports make that data easy to analyse.
3	Cision	Manage and track your brand's reputation by exploring what people are saying about your business across a variety of social platforms
4	People Browse	Find all the mentions of your brand, industry and competitors and analyse sentiment. This tool allows you to compare the volume of mentions before, during and after your marketing campaigns
5	Tweetstat	This is a fun, free tool that allows you to graph your Twitter stats. Simply enter your Twitter handle and "let the magic happen."
6	Steamcrab	Web application for sentiment analytics on Twitter data. It focuses on keyword searches and analyses tweets according to a two-pole scale (positive and negative). Visualization options are limited to scatter plots and pie charts.

4. Conclusion

The main goal of this paper is to evaluate the ensemble method for sentiment classification. This survey paper delivered an overview of recent updates in sentiment analysis and classification methods. Many of the articles cited in this paper give their contribution to the real-world application. Much

research has been carried on to mine the opinions in the contour of a document, sentence and feature level sentiment analysis. It has been examined that now the opinion mining trend is proceeding to the sentimental reviews of twitter data, comments used in Facebook on pictures, videos or Facebook status. Therefore, this paper discusses about an overview of the sentimental analysis approach of SA and OM in detail with the techniques and tools.

Reference

Aggarwal, R., & Gupta, L. (2017). A Hybrid Approach for Sentiment Analysis using

Classification Algorithm. *International Journal of Computer Science and Mobile*

Computing, *6*(6), 149–157.

Akter, S., & Aziz, M. T. (2017). Sentiment analysis on facebook group using lexicon

based approach. *2016 3rd International Conference on Electrical Engineering and*

Information and Communication Technology, ICEEiCT 2016, 8–11.

https://doi.org/10.1109/CEEICT.2016.7873080

Al-Hussaini, H., & Al-Dossari, H. (2017). A Lexicon-based Approach to Build

Service Provider Reputation from Arabic Tweets in Twitter. *International Journal of*

Advanced Computer Science and Applications, *8*(4), 445–454.

Ankitkumar, D., R. R. Badre, & Mayura Kinikar. (2014). A Survey on Sentiment

Analysis and Opinion Mining. *Proceedings of the International Conference on*

Advances in Information Communication Technology & Computing - AICTC '16, 1–5.

https://doi.org/10.1145/2979779.2979832

Anshuman, Rao, S., & Kakkar, M. (2017). A rating approach based on sentiment

analysis. *7th International Conference Confluence 2017 on Cloud Computing, Data*

Science and Engineering, 557–562.

https://doi.org/10.1109/confluence.2017.7943213

Augustyniak, L., Szymánski, P., Kajdanowicz, T., & Tuliglowicz, W. (2016).

Comprehensive study on lexicon-based ensemble classification sentiment analysis.

Entropy, *18*(1), 1–29. https://doi.org/10.3390/e18010004

Bzdok, D., Krzywinski, M., & Altman, N. (2018). Points of significance: Machine

learning: Supervised methods. *Nature Methods*, *15*(1), 5–6.

https://doi.org/10.1038/nmeth.4551

F.Y, O., J.E.T, A., O, A., J. O, H., O, O., & J, A. (2017). Supervised Machine

Learning Algorithms: Classification and Comparison. *International Journal of*

Computer Trends and Technology, *48*(3), 128–138.

https://doi.org/10.14445/22312803/IJCTT-V48P126

Gaur, M., & Pruthi, J. (2017). A Survey on Sentiment Analysis and Opinion mining.

International Journal of Current Engineering and Technology, *77*(22), 2277–4106.

https://doi.org/10.1145/2979779.2979832

Ghorbel, H., & Jacot, D. (2011). Sentiment analysis of french movie reviews. *Studies*

in Computational Intelligence, *361*, 97–108. https://doi.org/10.1007/978-3-642

21384-7_7

Gupta, P., Tiwari, R., & Robert, N. (2016). Sentiment Analysis and Text
Summarization of Online Reviews : A Survey. *International Conference on
Communication and Signal Processing*, 241–245.
https://doi.org/10.1109/ICCSP.2016.7754131

Inui, T., & Okumura, M. (2006). A Survey of Sentiment Analysis. *Journal of
Natural
Language Processing*, *13*(3), 201–241. https://doi.org/10.5715/jnlp.13.3_201

Jain, A. P., & Katkar, V. D. (2015). Sentiments analysis of Twitter data
using data
mining. *2015 International Conference on Information Processing (ICIP)*,
807–810.
https://doi.org/10.1109/INFOP.2015.7489492

Kinge, M., Nandedkar, P. S., & Narkhede, P. G. (2015). A Survey
onTrafficSentiment
Analysis. *International Journal of Advanced Research in Engineering
Technology &
Science (Ijarets)*, *2*(6), 1–7.

Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of
Classification
Techniques. *Informatica*, *31*, 249–268. https://doi.org/10.1115/1.1559160

Madhoushi, Z., Hamdan, A. R., & Zainudin, S. (2015). Sentiment analysis
techniques
in recent works. *Science and Information Conference, SAI 2015*, 288–
291.https://doi.org/10.1109/SAI.2015.7237157

P, C. A., & Patel, K. M. (2015). Sentiment Analysis Using Hybrid
Approach : A
Survey. *Journal of Engineering Research and Application*, *5*(1), 73–77.

Park, C. W., & Seo, D. R. (2018). Sentiment analysis of Twitter corpus
related to
artificial intelligence assistants. *2018 5th International Conference on
Industrial

Engineering and Applications, ICIEA 2018, 495–498.
https://doi.org/10.1109/IEA.2018.8387151

Pasarate, S., & Shedge, R. (2016). Comparative study of feature extraction
techniques

used in sentiment analysis. *2016 International Conference on Innovation and
Challenges in Cyber Security (ICICCS-INBUSH)*, (Iciccs), 182–186.
https://doi.org/10.1109/ICICCS.2016.7542328

Pradesh, U., & Pradesh, U. (2017). Sentiment Analysis on Product. *Internal
Conference on Computing, Communication and Automation (ICCCA 2017)*.

Priya, R. C. M., & Sathiaseelan, J. G. R. (2017). An Explorative Study on
Sentiment

Analy

sis. *2017 World Congress on Computing and Communication
Technologies(WCCCT)*,

140–142. https://doi.org/10.1109/WCCCT.2016.41

Rana, S., & Singh, A. (2016). Comparative Analysis of Sentiment
Orientation Using

SVM and Naïve Bayes Techniques. *International Conference on Next
Generation*

Computing Technologies, (October), 106–111.

Rezaei, Z., & Jalali, M. (2017). Sentiment analysis on Twitter using
McDiarmid tree

algorithm. *2017 7th International Conference on Computer and Knowledge
Engineering, ICCKE 2017, 2017–January*(Iccke), 33–36.
https://doi.org/10.1109/ICCKE.2017.8167924

Rizwana, K. H., Phil, M., Scholar, R., & Kalpana, B. (2018). A Survey On
Sentiment

Analysis And Opinion Mining. *International Journal of Pure and Applied
Mathematics (IJPAM)*, *118*(18), 2681–2688. Retrieved from
http://www.ijpam.eu

Routray, P., Kumar Swain, C., & Prava Mishra, S. (2013). A Survey on

Sentiment

Analysis. *International Journal of Computer Applications*, *76*(10), 975–
8887.

https://doi.org/10.5120/13280-0527

Salinca, A. (2016). Business Reviews Classification Using Sentiment
Analysis.

*Proceedings - 17th International Symposium on Symbolic and Numeric
Algorithms for*

Scientific Computing, SYNASC 2015, 247–250.

https://doi.org/10.1109/SYNASC.2015.46

Sarawgi, K. (2017). Opinion Mining : Aspect Level Sentiment Analysis
using

SentiWordNet and Amazon Web Services. *International Journal of
Computer*

Application, *158*(6), 31–36. https://doi.org/10.5120/ijca2017912830

Sathya, R., & Abraham, A. (2013). Comparison of Supervised and
Unsupervised

Learning Algorithms for Pattern Classification. *International Journal of
Advanced*

Research in Artificial Intelligence, *2*(2), 34–38.

https://doi.org/10.14569/IJARAI.2013.020206

Schouten, K., & Frasincar, F. (2016). Survey on Aspect-Level Sentiment
Analysis.

IEEE Transactions on Knowledge and Data Engineering, *28*(3), 813–830.

https://doi.org/10.1109/TKDE.2015.2485209

Sengottuvelan, P., & Karthik, A. V. (2017). Asurvey on Web and Rule
Based Traffic

Sentiment Analysis. *International Journal of Computer Trends and
Technology*,

43(1), 13–19. https://doi.org/10.14445/22312803/IJCTT-V43P104

Zhang, X., & Zheng, X. (2016). Comparison of Text Sentiment Analysis

Based on

Machine Learning. *2016 15th International Symposium on Parallel and Distributed*

Computing (ISPDC), 230–233. https://doi.org/10.1109/ISPDC.2016.39

Zvarevashe, K., & Olugbara, O. O. (2018). A framework for sentiment analysis with

opinion mining of hotel reviews. *2018 Conference on Information Communications*

Technology and Society, ICTAS 2018 - Proceedings, 1–4.

https://doi.org/10.1109/ICTAS.2018.8368746