

PAPER • OPEN ACCESS

A Review on Sentiment Analysis Techniques and Applications

To cite this article: Mohd Ridzwan Yaakub *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **551** 012070

View the [article online](#) for updates and enhancements.

A Review on Sentiment Analysis Techniques and Applications

Mohd Ridzwan Yaakub¹, Muhammad Iqbal Abu Latiffi², Liyana Safra Zaabar³

^{1,2,3}Center for Artificial Intelligence Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia.

E-mail: ridzwanyaakub@ukm.edu.my

Abstract. Nowadays, what user think is the most difficult and complicated task handled by organizations. The way to identify the attitude of the speaker or a writer on some topics is to use sentiment analysis. The use of sentiment analysis is to identify user's opinion towards some topics whether it is positive or negative. This paper presents the techniques used by previous researchers in sentiment analysis which are Machine Learning and Natural Language Processing (NLP) in solving the classification task. The comparison among these two main approaches reveals that Machine Learning techniques can solve classification task with reasonable success and with very high accuracy compared to NLP-based techniques but it is depending on the training and test data with respect to the domain. This paper also presents the use of ontology in sentiment analysis that can help in achieving more high accuracy for the classification task.

1. Introduction

Nowadays, sharing opinion, sharing marketing, online ticket bookings, and online shopping are rising in daily life of people. In the decision-making process, we are always concerned about what other people think. Other people's opinions are the most important piece of information for most of us. Social networking sites such as Facebook allow users to post their comments or opinions about any issues and topics. Here, we will see the importance of reviews shared by the peoples.

Sentiment can be defined as a view or opinion that is held or expressed. Whereas, sentiment analysis (SA) is a process of computationally identifying and categorizing opinions expressed in a piece of textual content, particularly to decide the writer's attitude towards a particular topic, product or issue [1]. SA also known as opinion mining is widely used in many domains such as products, services, issues and politics to analyze user's behaviors or opinion regarding the related topics [2].

People's opinions or experiences influence users in making a decision. Typically, the use of SA is to extract what people feel or think about different things from the text. In the business industries, the use of SA is to analyze customer's opinions towards their product or services because user's satisfaction is the main attention in their industries. SA technology helps in improving businesses by analyzing user's opinion.

2. Techniques Used in Sentiment Analysis

SA aim is to identify the positive, negative or neutral sentiment in the set of text or documents. It is a very crucial task to build a system to analyze user's opinion in the form of review, social media, and microblogs. A lot of researches on SA have been done [3-5]. There are two main approaches used in SA which are Natural Language Processing (NLP) based and Machine Learning based.



NLP is commonly used for text mining, machine interpretation and mechanized inquiry. According to [6], NLP is a research and application area that explores how computers can be used to understand and manipulate natural language text or language for useful purposes. Furthermore, NLP researchers focus on trying to gather knowledge about how people understand and use language so that relevant mechanisms and techniques can be developed in allow computer systems to know as well as could manipulate natural languages to accomplish desired tasks. NLP method is broke down into basic methods and methods of modelling. The basic method in the NLP technique is to extract text features such as Part-of-Speech (POS) Tagging, frequency of documents, dictionary and weighting. In the meantime, the topic modelling method is a generic model of probability using distribution of vocabulary in searching for topics with text elements [7]. Both of these methods aim to identify the topic from a collection of documents. Based on the previous study, NLP approach uses variety of techniques in a different type of domain and most of the result is above 80% which can be say it is a quite impressive compared to Machine Learning approach. NLP approach had proven successful in [6,8-11] to get an excellent result.

Typically, NLP methods utilize statistical algorithms, which will eventually base or combine the algorithms with a classical Deep Learning Approach. Research in NLP is essential for opinion mining as it can monitor positive or negative human feelings based on the sentence or feature extraction in a user's comment.

Another method used in sentiment analysis is Machine Learning. Machine Learning is a set of statistical technique for analyze some form of text. The techniques can be indicate as a model that is then applied to other text (supervised), or could be a set of algorithms that work across large sets of data to extract meaning (unsupervised) [12]. Machine Learning approach has various of techniques, but there are a few techniques that most of the researchers frequently used.

Generally, among the popular Machine Learning methods, Support Vector Machine (SVM), Naïve Bayes and N-Gram are the most popular. SVM is a discriminative classifier formally defined by a separating hyperplane. In other words, having given labelled training data (supervised learning), the algorithm from SVM outputs an optimal hyperplane which categorizes new examples [13]. According to [14] The Naïve Bayesian classifier is based on Bayes' theorem with independence assumptions between predictors, and is easy to build, with no complicated iterative parameter estimation which makes it particularly useful for very large datasets. [14] also stated that despite its simplicity, the Naïve Bayes classifier often does surprisingly well and is widely used because it often outperforms more sophisticated classification methods. Additionally, N-gram is a simple model that assigns probabilities to sentences of words or of whole sequences. N-gram is widely used in text mining and NLP. It is one of the most important tools in speech and language processing. N-gram has been used for a variety of different tasks. These three techniques are the most used in SA because it is proven to achieve high accuracy in the previous researches such as in [15-24]. Machine learning method is able to eliminate overlapping and irrelevant features by using suitable feature selection method. However, according to [5], Machine Learning methods suffers a great deal when training on mixed data.

Machine Learning approach are the best technique among two approaches and able to achieve very excellent results. [25] is among the earlier research in SA that used machine learning approach in solving the classification task and is able to achieve very excellent results. Although this paper is in an early research, they managed to obtain the most satisfactory result up to 96% of accuracy compare to NLP approach where the highest accuracy is over 80% by [10]. Machine learning approach precedes NLP-based approach by 10% and based on these results, we can simply say that machine learning is the best method compared to NLP-based techniques according to the classification task.

3. Ontology in Sentiment Analysis

Ontologies are the key component in various fields such as information retrieval and extraction, knowledge management and organization. Ontologies has been used for modeling the terms in a domain of interest as well as the relations among these terms and are now applied in many fields [26]. The study by [25], is one of the early research in SA that use ontology in sentiment classification. In

this paper they presents a method of ontology-based sentiment classification to classify and analyze online product reviews of consumers. In this research, the author used SVM text classification approach based on a lexical variation ontology. The use of ontology in this paper helped the classifier capture and analyzed the sentiments that use different words which may share the same meaning. The performance measure using precision, recall and F-measure allowed the author to be able classify with good accuracy after testing by F-measure with 96% accuracy. Many works have been done in ontology deployment in SA. Other than [25], researches by [27-31] are among the current researches in SA that uses ontology and managed to get a very satisfying results. The results in every experiment proved that the used of ontology can perform better than all baseline.

4. Conclusions

SA present a important role in decision making. Organizations will make decision about their product or service based on the reviews. There are a lot of approaches that can be used in SA. The Machine Learning techniques can solve classification task with reasonable success and with a very high accuracy compared to NLP- based techniques. From these results, we can conclude that even though Machine Learning obtained the highest results, but it is depending on the training and test data with respect to domain. This paper also presented the use of ontology in sentiment analysis which can obtain a very high accuracy.

Acknowledgements

The authors would like to thank Center for Artificial Intelligence Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia by giving the authors an opportunity to conduct this research and funded by Universiti Kebangsaan Malaysia and Malaysia Government under the FRGS research grant scheme with code FRGS/1/2017/ICT02/UKM/02/4.

References

- [1] Medhat W *et al* 2014 *Ain Shams Eng. J.* **5**(4) 1093–1113
- [2] Kumar P N V S P *et al* 2018 *Proc. 2nd Int'l Conf. on Comp. Intelligence and Informatics* **712**
- [3] Pang B *et al* 2002 *Proc. Conf. Empir. Methods Nat. Lang. Process.* 79–86
- [4] Gamon M 2004 *Proc. 20th Int'l Conf. Comput. Linguist* **January 2004** 841–es
- [5] Read J 2005 *ACL Student Res. Work.* **June** 43–48
- [6] Jurek A *et al* 2015 *Secur. Inform.* **4**(1) 9
- [7] Kim H *et al* 2011 *Comprehensive Review of Opinion Summarization* (Illinois Dig. Env. ...)
- [8] Khan A *et al* 2010 *Proc. 8th Int'l Conf. Front. Inf. Technol.* 1–6
- [9] Palanisamy P *et al* 2013 *Proc – Int Workshop on Semantic Evaluation* 543–548
- [10] Vu L 2017 *Int'l Conf. Inf. Knowl. Eng.* (July) 10–16
- [11] Feng J *et al* 2018 *Wirel. Commun. Mob. Comput.* **2018**
- [12] Abirami A *et al* 2016 *IEEE Eighth Int. Conf. Adv. Comput.* 72–76
- [13] OpenCV 2017 *Introduction to Support Vector Machines — OpenCV 2*
- [14] Sayad D S 2010 *Naive Bayesian* 1–4
- [15] Dhaoui C *et al* 2017 *J. Consum. Mark.* **34**(6) 480–488
- [16] Go A *et al* 2009 *Processing* **150**(12) 1–6
- [17] Tripathy A *et al* 2015 *Procedia Comput. Sci.* **57** 821–829
- [18] Xia R *et al* 2016 *Inf. Process. Manag.* **52**(1) 36–45
- [19] Trivedi S K *et al* 2018 *Electron. Libr.* **36**(4) 590–606
- [20] Al-Smadi M *et al* 2018 *J. Comput. Sci.* **27** 386–393
- [21] Devika M D *et al* 2016 *IOSR J. Comput. Eng.* 1–5, 2016.
- [22] Rana S *et al* 2017 *Proc. 2016 2nd Int. Conf. Next Gener. Comput. Technol.* **October** 106–111
- [23] Ibrahim M N M *et al* 2018 *2017 IEEE Conf. Open Syst.* **2018–January** 17–20
- [24] Zvarevashe K *et al* 2018 *2018 Conf. Inf. Commun. Technol. Soc. - Proc.* 1–4
- [25] Polpinij J *et al* 2008 *Proc. - 2008 IEEE/WIC/ACM Int. Conf. Web Intell. WI 2008* 518–524

- [26] Gomez-Perez A *et al* 2002 *IEEE Intell. Syst.* **17**(1) 54–60
- [27] Liu L *et al* 2012 *2012 5th Int. Congr. Image Signal Process.* (Cisp) 1620–1624
- [28] Yaakub M R *et al* 2013 *Procedia Technol.* **11**(June 2015) 495–501
- [29] Sam K M 2013 *Int. J. e-Education, e-Business, e-Management e-Learning* **3**(6)
- [30] Schouten K *et al* 2017 *Proc - Int. Conf Web Eng* 302-320
- [31] Ali F *et al* 2016 *Appl. Soft Comput. J.* **47**(November 2018) 235–250