Discovering Hate Sentiment within Twitter Data through Aspect-Based Sentiment Analysis

To cite this article: Nurulhuda Zainuddin et al 2020 J. Phys.: Conf. Ser. 1447 012056

View the article online for updates and enhancements.
Discovering Hate Sentiment within Twitter Data through Aspect-Based Sentiment Analysis

Nurulhuda Zainuddin¹,², Ali Selamat¹,², Roliana Ibrahim²
¹Malaysia Japan International Institute of Technology (MJIT), Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, Kuala Lumpur, Malaysia
²Media and Games Center of Excellence (MaGiCX) UTM & School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Malaysia
³Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Melaka, 77300 Merlimau, Melaka, Malaysia

E-mail: aselamat@utm.my

Abstract. Aspect-based sentiment analysis is a vital issue in fine-grained sentiment evaluation, which intends to provide an automatic prediction of the sentiment polarity, given a particular aspect in its context. This paper presents an aspect-based sentiment analysis to find hate sentiment inside twitter data. Word embeddings have had prevalent utilisation in Natural Language Processing (NLP) applications because their vector representations have the ability to capture useful linguistic relationships and semantic properties between words with the help of deep neural networks. Word embeddings have often been used in machine learning models as feature input, which allows for the contextualisation of raw text data in machine learning techniques. The model has the ability to represent the relationship between the word embedding features and the aspects as feature representation within the suggested model. To assess the efficacy of the proposed method, extensive experiments were performed on the dataset of the researcher, as well as on widely utilised datasets. It was demonstrated by the experimental results that the proposed method was able to obtain impressive results among the three datasets.

1. Introduction
Twitter is a social networking and online news service that holds a large quantity of information. ‘Tweets’, which are the messages posted on Twitter, have a 150-character limit. Millions of people use Twitter, and the site is known as one of the most visited sites throughout the world. On average, there are 58 million tweets posted every day [1]. Individuals and organisations have also increased their social media content from different areas, including Facebook, community forums, online reviews, and blogs to help in making decisions. There is a lesser need for people to depend on the opinions of their friends and family members because users post numerous reviews on the internet about various products and services. Tweets can exist as either messages or images that users on social networking sites post. These tweets can offer individuals and organisations valuable information.

Over the past several years, there has been a significant rise in the quantity of interactive contents on the Internet. Websites now allow users to post whatever they want. However, the rise of social networks like Twitter, Facebook and Google+ has also been the root of many social phenomena such as intimidation, harassment, depression, hate speech, and suicide. People have a tendency to express their biases or their feelings towards specific groups of people. It is difficult to identify hate speech in social media because in such environments, the textual contents are often informal, unstructured, and the words are often spelled wrong. There has been increasing interest in such tasks, along with the challenges that come along with them, such as the amount of data that need processing and the complexity of handling unstructured text that is expressed in natural languages. As a result, many fields have been influenced, including sentiment analysis (SA) and big data.

The murder of Drummer Lee Rigby in Woolwich in 2013 resulted in a widespread social media reaction. Given the public nature of the actions and the extreme terrorist motive, there was a possibility that the public response may include written expressions of antagonistic and hateful
sentiment towards a specific ethnicity, race, and religion, which one can interpret as ‘hate speech’. From the hate speech sample, it was apparent most of the terms utilised in hate speech were derogatory or expletive, targeted at particular social groups such as religion, race, disability, ethnic origin, nationality, or sexual orientation [4, 5]. A research by [6] conducted a study that captures the trend of rising anti-Korean sentiment and hate speech in present Japan. The study also determined the on-going history of the battle against those who attempt to ruin society in Japan by base-hearted means. The study presented a short list of remarks and street propaganda that Japanese racists expressed during demonstrations such as “If you hate Japan, why don’t you go home?” or “You can throw stones at Korean. You can rape Korean ladies” [6]. Furthermore, the assault of a person because he is Black constitutes a hate crime, while assaulting a Black person for other reasons will be classified as simple assault. Certain states have put in place specific “hate crime” offences while others increase the severity of the primary charge or increase the sentence when it is determined that hate has been motivating the crime. This study makes use of aspect-based sentiment analysis (ABSA) for the analysis of Twitter contents that are regarded as crime. Unlike traditional sentiment analysis, ABSA also makes use of several added challenges such as: (a) connecting every portion of the text with the aspect it’s referring to, (b) determining the parts of text that talk about the same aspect [7]. The aim of this study is to tackle these problems by offering an improved semantic model for the aspect-based sentiment analysis (ABSA) method. The suggested approach comprises: (a) baseline evaluation method and research to support the comparison of the suggested study using the researcher’s dataset and the reference dataset, and (b) an enhanced twitter ABSA that has its basis on word embedding features.

The remainder of this paper is organised as follows. Section 2 discusses an associated work for sentiment analysis and aspect-based sentiment analysis. Section 3 will present a model architecture comprising the features utilised in the experiment and discussion on the baseline approach and assessment. Section 4 presents the results and analysis. Section 5 provides the conclusions and suggestions for future research.

2. Related Works on Aspect-Based Sentiment Classification

Although it is useful to classify opinionated texts at the sentence or document level, it does offer the necessary information that is required for several other applications. A positive opinionated document about a specific entity does not automatically signify that the author possesses positive opinions on every aspect of that entity. Similarly, a negative opinionated document is not an indication that the author dislikes everything. Given a typical opinionated document, both negative and positive qualities of the entity are written by the author, although the entity may have a positive or negative general sentiment.

Sentence and document sentiment classification do not offer such information. These details are obtained by going to the aspect level. In other words, the full model of aspect-based sentiment analysis is needed. Instead of viewing opinion mining simply as a way to classify sentiments, aspect-based sentiment analysis presents a collection of problems that need deeper natural language processing capabilities to generate a more enriched set of results. The task contains four subtasks: aspect term extraction, aspect category identification, aspect-terms’ polarity identification, and aspect category polarity.

Aspect-based sentiment classification is used in classifying the opinions on various aspects as either subjective or objective. The primary task in aspect-based sentiment classification is to find the opinion words and then determine the polarity of these words towards their context. One of the methods that researchers use involves extracting all the adverbs, adjectives, and verbs from the tagged sentences [11, 13, 14, 15]. In this method, [12] utilised linguistic rules for the determination of word opinion orientation. In this process, the set of adjectives that are seen near each aspect are also extracted, instead of just utilising the linguistic rule.

However, other alternatives exist to transform text to numbers. Generally, the term word embedding is used to refer to the techniques that make it possible to represent the words of a certain
vocabulary over a continuous vector space of a dimension which is significantly lower than the vocabulary size. These techniques are able to extract a word profile in an unsupervised way by simply considering the contexts in which these words are seen. For instances, one research [2] addresses on how to integrate different representations of input for the problem of aspect-based sentiment analysis. They also focus on three kinds of representation including word embeddings from the two methods (Word2Vec and GloVe) and the one-hot character vectors. Besides, [3] proposes to utilize both word sentiment label from lexicon and tweet sentiment label from distant supervised information for training sentiment-enriched word embedding. The method integrates word-level information into tweet-level as a convolutional result, while simultaneously modelling word-level n-gram and sentiment information.

Word embedding techniques are seen to bridge the lexical frontier because word representation also allows for the capturing of semantic and syntactic aspects. Word2Vec [16] and GloVe [27] are two recognised and accepted methodologies that are used for encoding words as real-value vectors. Word2Vec constitutes of an “artificial neural network predictive” model that includes the continuous bag of words (CBOW) and Skip-gram architectures. On the other hand, GloVe is a “count based” model. In the CBOW model, the current word is envisaged from its context words. On the contrary, the Skip-gram model makes use of the current word to predict its context words. Alternatively, the GloVe model operates on an aggregated global word–word co-occurrence statistics from a corpus. Hence, the two approaches, Word2Vec and GloVe, have the potential to encode different aspects of a language. In addition to this, if different corpora are used, then it will apparently create varying word vectors where every vector is composed of highly complementary information with respect to the others.

Google developed Word2vec in 2013, which emerged to be a powerful tool for text processing [16]. Word2Vec is one of the words embedding tools that have recently gained more attention within the scientific community. This model uses neural networks as its basis and has a relationship to some of the elements that can be credited for the advent of the so-called deep learning. In high-dimensional vector space, it exhibits high efficiency when it comes to computing word vector representations. In the vector space, word vectors emerge from words that have comparable semantic and share common contexts and are mapped close to each other. Besides syntactic information, the similarity of word representations based on semantic characteristics such as semantic relationships are usually conserved in vector operations carried out on word vectors. In this study, the Word2Vec method is used as it has been established that Word2Vec creates better word embeddings for most general NLP tasks compared to other approaches.

3.0 Model Architecture
The aggregated tweets undergo distribution with regards to sentiment polarity. For this segment of ABSA, the predicted label tweets are being classified as either negative or positive, while the neutral category is not taken into consideration. For the entire dataset, majority of the tweets were in the negative category, while the positive category had a significantly small number of tweets. The key cause behind this trend is that most of the Twitter users use Twitter to voice out their complaints and channelize their negative energy arising from failed service and negative experiences. On the contrary, very few use it for sharing pleasant experiences while using products or services, or for positive commentaries that emphasise their support for the target group. The model pre-processes tweets from the researcher’s HCTS dataset that includes 1,078 tweets [17]. Additional datasets that were used include the Stanford Twitter Sentiment (STS) dataset and the Sanders Twitter Corpus (STC) dataset [18].

This paper focuses on assessing the usefulness of the representations given by Word2Vec in classifying documents. Our primary goal is to determine if the knowledge Word2Vec offers about words is complementary to the information that is gathered using a very stable model for the task of
classifying. In this case, Word2Vec has an even more interesting contribution, as it demonstrates the utility of a more semantic representation given a more demanding situation.

3.1 Features used for the experiment

N-grams [10, 21, 22] and lexicon characteristics [8, 9, 19, 20] are used for some of the researches. Similarly, Twitter has its unique set of attributes and features. This may accordingly impact some of the researches that use hash tags as parameters for analysis [23, 24]. Moreover, other significant characteristics of Twitter include the keyword, word context and statistical features [25]. In this study, the following characteristics are used for analysis:

(i) Word2Vec was used to signify words in the neural network training. The classification of words was conducted using feature representation. Each word was encoded as a 300-dimensional vector, which was then supplied to the network. It was revealed by the researchers that the use of word vectors that have been trained in this manner exhibit significant efficiency with respect to semantic similarity comparison.

(ii) POS was used that represents verbs, adverbs, nouns, and adjectives.

(iii) N-grams were used for feature representation in the classification. In the text classification, unigrams are commonly used as features, which are nothing but single words. These basic unigrams are highly successful for feature representation in sentiment classification [28]. Only those words that appear more than once in the tweet’s dataset are used. The unigrams are derived, and their frequency is counted, and if the frequency is found to be greater than one, the word is included as a feature and the numerical value is set as its frequency.

4.0 Results and Analysis

Table 1 presents the results with respect to the terms of the percentage of recall and precision that are derived through three versions of domain-specific classifiers. It is found that the word embeddings trained on HCTS exhibit superior traits compared to the other features when all the metrics (precision: 83.3, recall: 75, F1 score: 73.3) are taken into consideration. This is in contrast to the F-measure of 61.8 for the unigram features, and 53.3 for the POS tags features. The performance for STC dataset is better when Word2Vec is used with F-measures 71.4%, compared to 71.3% for POS tags features and 64.5% for unigram features. Moreover, the performance is comparable for the STS Dataset, where it achieves 92.9% with Word2Vec features, and 86.2% with POS tags and unigram features. It is observed that the three-domain dataset performance with Word2Vec features leads to improved performance in relation to the use of other features. Although, in all the cases, recall was considerably lower than precision.

5.0 Conclusion and future works

In this study, an aspect-based sentiment analysis is carried out to analyse and assess the prevalence of hate sentiment within Twitter data. In the proposed model, it is possible to represent the relationship between the word embedding features and the aspects as feature representation. A set of classifiers were implemented in association with different features, and then compared in terms of their performance on the ABSA tasks. In addition to this, the word embedding features were used to play a major role in improving the prediction results. The future plan is to integrate more information into this model, like additional features such as topic to improve the capability of Twitter word embedding. On the other hand, the plan also involves the evaluation of the recommended approaches on other social media datasets such as Facebook, LinkedIn, YouTube, etc. The important aspect to consider is that the results will offer a direction to future researches that might be beneficial for aspect-based sentiment classification performance.

Table 1: Comparison of methods

| Dataset   | Classifier | Precision | Recall | F-measure | Precision | Recall | F-measure | Precision | Recall | F-measure |
|-----------|------------|-----------|--------|-----------|-----------|--------|-----------|-----------|--------|-----------|
|           |            |           |        |           |           |        |           |           |        |           |
### Acknowledgements

The research is supported by Universiti Teknologi Malaysia (UTM) under Research University Grant Vot-20H04, Malaysia Research University Network (MRUN) Vot 4L876 and the Fundamental Research Grant Scheme (FRGS) Vot 5F073 supported under Ministry of Education Malaysia.

### References

[1] Number of monthly active international twitter users from 1st quarter 2010 to 1st quarter 2018 (in millions), https://www.statista.com/statistics/274565/monthly-active-international-twitter/

[2] D.-H. Pham, A.-C. Le, Exploiting multiple word embeddings and one-hot character vectors for aspect-based sentiment analysis, International Journal of Approximate Reasoning 103 (2018) 1 – 10.

[3] S. Xiong, H. Lv, W. Zhao, D. Ji, Towards twitter sentiment classification by multi-level sentiment-enriched word embeddings, Neurocomputing 275 (2018) 2459 – 2466.

[4] F. B. of Investigation, Training guide for hate crime data collection: Summary reporting system, national incident-based reporting system., Tech. rep., Washington, DC (1992). URL http://www.fbi.gov/about-us/investigate/civilrights/hatecrimes/overview

[5] R. Stotzer, Comparison of hate crime rates across protected and unprotected groups (June 2007). URL http://williamsinstitute.law.ucla.edu/wp-content/uploads/Stotzer-Comparison

[6] K. Ito, Anti-korean sentiment and hate speech in the current japan: A report from the street, Procedia Environmental Sciences 20 (0) (2014) 434 – 443, the 4th International Conference on Sustainable Future for Human Security {SUSTAIN} 2013.

[7] M. Al-Smadi, M. Al-Ayyoub, Y. Jararweh, O. Qawasmeh, Enhancing aspect-based sentiment analysis of arabic hotels’ reviews using morphological, syntactic and semantic features, Information Processing & Management, 2019, 56(2), 308-319.

[8] V. Khuc, C. Shivade, R. Rammath, J. Ramanathan, Towards building large-scale distributed systems for twitter sentiment analysis, 2012, pp. 459–464.

[9] G. Paltoglou, M. Thelwall, Twitter, myspace, digg: Unsupervised sentiment analysis in social media, ACM Transactions on Intelligent Systems and Technology, 2012, 3 (4).

[10] Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, Sentiment analysis of twitter data, in: Proceedings of the Workshop on Languages in Social Media, LSM ’11, Association for Computational Linguistics, Stroudsburg, PA, USA, 2011, pp. 30–38.

[11] W. Zhang, H. Xu, W. Wan, Weakness finder: Find product weakness from chinese reviews by using aspects-based sentiment analysis, Expert Systems with Applications 39 (11) (2012) 10283 – 10291.

[12] E. Marrese-Taylor, J. D. Velquez, F. Bravo-Marquez, A novel deterministic approach for aspect-based opinion mining in tourism products reviews, Expert Systems with Applications 41 (17) (2014) 7764 – 7775.

[13] H. H. Lek, D. Poo, Aspect-based twitter sentiment classification, in: Tools with Artificial Intelligence (ICTAI), 2013 IEEE 25th International Conference on, 2013, pp. 366–373.
[14] J. Jmal, R. Faiz, Customer review summarization approach using twitter and sentiwordnet, in: Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics, WIMS ’13, ACM, New York, NY, USA, 2013, pp. 33:1–33:8.

[15] H. Kansal, D. Toshniwal, Aspect based summarization of context dependent opinion words, Procedia Computer Science 35 (0) (2014) 166 – 175, knowledge-Based and Intelligent Information &amp; Engineering Systems 18th Annual Conference, KES-2014 Gdynia, Poland, September 2014 Proceedings.

[16] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, CoRR, 2013. abs/1301.3781. arXiv:1301.3781. URL http://arxiv.org/abs/1301.3781

[17] N. Zainuddin, A. Selamat, R. Ibrahim, Hybrid sentiment classification on twitter aspect-based sentiment analysis, Applied Intelligence, 2018, 48(5), 1218-1232.

[18] H. Saif, M. Fernandez, Y. He, H. Alani, Evaluation datasets for twitter sentiment analysis: A survey and a new dataset, the sts-gold, 2013.

[19] F. Bravo-Marquez, M. Mendoza, B. Poblete, Combining strengths, emotions and polarities for boosting twitter sentiment analysis, in: Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining, WISDOM ’13, ACM, New York, NY, USA, 2013, pp. 2:1–2:9.

[20] E. Baucom, A. Sanjari, X. Liu, M. Chen, Mirroring the real world in social media: Twitter, geolocation, and sentiment analysis, in: Proceedings of the 2013 International Workshop on Mining Unstructured Big Data Using Natural Language Processing, UnstructureNLP ’13, ACM, New York, NY, USA, 2013, pp. 61–68.

[21] H. Wang, D. Can, A. Kazemzadeh, F. Bar, S. Narayanan, A system for real-time twitter sentiment analysis of 2012 U.S. presidential election cycle, in: Proceedings of the ACL 2012 System Demonstrations, ACL ’12, Association for Computational Linguistics, Stroudsburg, PA, USA, 2012, pp. 115–120.

[22] Marcus, M. S. Bernstein, O. Badar, D. R. Karger, S. Madden, R. C. Miller, Twitinfo: Aggregating and visualizing microblogs for event exploration, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI ’11, ACM, New York, NY, USA, 2011, pp. 227–236.

[23] X. Wang, F. Wei, X. Liu, M. Zhou, M. Zhang, Topic sentiment analysis in twitter: A graph-based hashtag sentiment classification approach, in: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, CIKM ’11, ACM, New York, NY, USA, 2011, pp. 1031–1040.

[24] E. Otsuka, S. A. Wallace, D. Chiu, Design and evaluation of a twitter hashtag recommendation system, in: Proceedings of the 18th International Database Engineering & Applications Symposium, IDEAS ’14, ACM, New York, NY, USA, 2014, pp. 330–333.

[25] T. Sakaki, M. Okazaki, Y. Matsuo, Earthquake shakes twitter users: Real-time event detection by social sensors, in: Proceedings of the 19th International Conference on World Wide Web, WWW ’10, ACM, New York, NY, USA, 2010, pp. 851–860.

[26] F. Sebastiani, Machine learning in automated text categorization, ACM Comput. Surv. 34 (1) (2002) 1–47.

[27] J. Pennington, R. Socher, C. D. Manning, Glove: Global vectors for word representation, in: A. Moschitti, B. Pang, W. Daelemans (Eds.), Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL, ACL, 2014, pp. 1532–1543.

[28] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up? Sentiment classification using machine learning techniques, in: Proceedings of the ACL02 Conference on Empirical Methods in Natural Language Processing - Volume 10, EMNLP ’02, Association for Computational Linguistics, Stroudsburg, PA, USA, 2002, pp. 79–86.