

Word2Vec model for sentiment analysis of product reviews in Indonesian language

M. Ali Fauzi

Faculty of Computer Science, Brawijaya University, Indonesia

Article Info

Article history:

Received Feb 3, 2018

Revised Jun 22, 2018

Accepted Jul 6, 2018

Keywords:

Sentiment analysis

Support vector machine

Text classification

Word embedding

Word2Vec

ABSTRACT

Online product reviews have become a source of greatly valuable information for consumers in making purchase decisions and producers to improve their product and marketing strategies. However, it becomes more and more difficult for people to understand and evaluate what the general opinion about a particular product in manual way since the number of reviews available increases. Hence, the automatic way is preferred. One of the most popular techniques is using machine learning approach such as Support Vector Machine (SVM). In this study, we explore the use of Word2Vec model as features in the SVM based sentiment analysis of product reviews in Indonesian language. The experiment result show that SVM can performs well on the sentiment classification task using any model used. However, the Word2vec model has the lowest accuracy (only 0.70), compared to other baseline method including Bag of Words model using Binary TF, Raw TF, and TF.IDF. This is because only small dataset used to train the Word2Vec model. Word2Vec need large examples to learn the word representation and place similar words into closer position.

Copyright © 2019 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

M. Ali Fauzi,

Faculty of Computer Science, Brawijaya University,

Jl. Veteran, Malang, Indonesia.

Email: mocha.ali.fauzi@ub.ac.id

1. INTRODUCTION

Since the rise of Web 2.0, the internet has become more user centric [1]. People are participating in making more and more content on the Internet through social media, discussion boards, Web forums, and blogs. Concurrently with such trends, an increasing number of websites where consumers can write and read reviews, and express their experiences, feeling, opinions, views, and complaints about various products and services has emerged [2]. From a consumer behavior perspective, it can be called as one of the greatest developments on the Internet.

both consumers and producers. In making purchase decisions, consumers often seek advice and purchase recommendations from others [3-4]. Previously, consumers commonly refer to advertisements in mass media to make this decision [5]. However, with the growth of e-commerce and increasing number of online review platforms, online reviews have become a reference for consumers they can rely on in finding information about the product to be purchased [6-7]. Consumers tend to learn how others like or dislike a product before buying. In fact, previous research found that consumers believe that online reviews provided by other users are more credible and trustworthy than the traditional sources [8].

For producer, online reviews can become a reference about what people think about their products or services to predict public acceptance level of their products. This information can help to forecast product sales. Furthermore, negative reviews can be the basis in product improvement and marketing strategies [9].

Therefore, understanding such sentiment and opinion information has become more and more prominent for both producers and customers. However, it becomes more and more difficult for people to understand and evaluate what the general opinion about a particular product in manual way since the number of reviews available increases. Hence, the automatic way is preferred.

Sentiment analysis, also known as sentiment or polarity classification, is a work of analyzing people's opinion or sentiment from a piece of text - for example to decide whether the sentiment is positive or negative [10]. We can consider sentiment analysis as text classification problem with sentiment as its classes. Nevertheless, sentiment classification is more challenging than traditional topic-based classification due to the necessity to extract more implicit information, instead of only keywords [11].

One of the most popular techniques is using machine learning approach. In recent years, sentiment classification using machine learning methods have been widely adopted and proven to provide supreme performance [12-17]. Prior research conducted by [10] also showed that machine learning techniques have quite good performance with SVMs tend to do the best. Two key issues in machine learning approach are how to extract complex features and finding out which kinds of features are more valuable [18]. Several feature extraction methods have been proposed such as single words [19-20], n-grams [21-22], lexicon [23], textual features [24], and many other new models [25-27]. However, semantic features have been infrequently employed in this field. Semantic features can disclose the implicit semantic relationships between words, which is should be useful for improving the sentiment classification performance.

Word embedding, also known as distributed word representation [28], is feature learning technique in Natural Language Processing (NLP) where words from the vocabulary are represented to low-dimensional vectors of real numbers [29]. By using word embedding, the semantic and syntactic information of words can be captured from a large number of unlabeled corpora [30-31]. Word embedding have been employed in many works in Natural Language Processing (NLP) to produce more effective word representations [32-36]. One of the most popular example of word embedding is Word2Vec model. Word2Vec [37] maps each words in the vocabulary into a dense vectors of real numbers using a shallow neural probabilistic language model [38]. By using Word2vec, words that similar will be close to each other in the embedding space [39].

In this study, we will explore the use of Word2Vec model for sentiment analysis of product reviews in Indonesian language. Word2Vec will be used as feature representation. For the classification task, we will use Support Vector Machine due its supreme performance. We will also explore the use of Bag of Word (BOW) model utilizing several term weighting methods including Term Frequency (TF) and Term Frequency-Inverse Document Frequency (TF.IDF).

2. RESEARCH METHOD

The general flowchart of the sentiment analysis system in this study is shown in Figure 1. There are three main stages in this system i.e. preprocessing, building Word2Vec model and classification using SVM. Each review will be classified into positive or negative class.

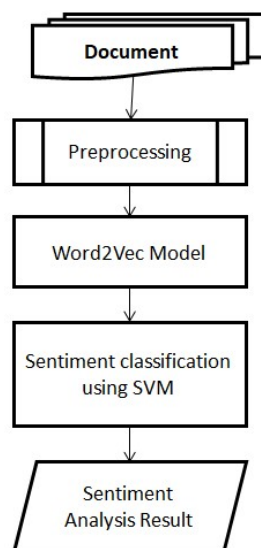


Figure 1. System main flowchart

2.1. Preprocessing

Preprocessing is conducted before the main process begin. Some steps conducted in this stage including tokenization, case folding and cleaning [40-43]. In tokenization, each review is splitted into smaller units called tokens or terms [44]. Case folding is a task of converting all of characters in review text become lowercase [45]. Meanwhile, in cleaning, characters outside of the alphabet such as punctuation, numbers, and html tag is omitted. In this study, stemming and filtering are not conducted because in some previous studies, stemming and filtering cannot improve sentiment analysis performance.

2.2. Building Word2Vec model

After the preprocessing stage was done, we build word vector representation using Word2Vec. First, the Word2Vec model builds a vocabulary from training data. Then, it learns and determines the vector representation of each words. There are two training algorithms in word2vec, i.e. continuous bag-of-words (CBOW) and skip-gram [46]. In this study, CBOW is employed. In CBOW, the word vector is built by predicting each word coocurrence based on its neighboring words. The resulting word vector will be employed as the classification features. Word2Vec generally can help to improve classification performance because in Word2Vec, the similar words have similar vectors.

2.3. Sentiment classification using support vector model

Finally, in the last stage, the reviews are classified into positive or negative class. In this study, support vector machines (SVMs) is used for the classification task. Despite its high computational complexity [47], SVM has become a popular algorithm in the last decade because of its excellent performance in text classification field [48].

Based on the representation of training data in feature space, SVM finds a hyperplane that separates the positive and negative data with maximum margin. Then, the testing data are then mapped into that same feature space and predicted to belong to positive or negative category based on which side they fall. In this study, we use linear kernel because based on the work of Mc Callum and Nigam [49], linear SVM has the best performance in text classification. The other benefit of linear kernel is that it is faster and require fewer parameters than other kernels in SVM.

3. RESULTS AND ANALYSIS

Experiment is conducted by using 772 product reviews extracted from FemaleDaily website. The text reviews and their ratings were collected and labelled manually from the website (<https://femaledaily.com/>). There are 386 reviews labelled as positive and 386 reviews labelled as negative. All of the reviews is in Indonesian language. Scikit-Learn [50] was used to implement the experiments. In the experiments, we compared the results of sentiment classification using Word2Vec with the other methods including Bag of Words (BOW) using Binary TF, Raw TF, and TF.IDF. We use 10-fold cross validation, which means the product reviews dataset is equally divided into 10 folds. We iterate the experiment 10 times. In each iteration, reviews from 9 folds were used as training data and the remaining one-fold was used as testing data. Average accuracy was used as the evaluation method. Experiment results can be seen in Figure 2.

Figure 2 depict that sentiment analysis using SVM generally have good performance with average classification accuracy value 0.81. The best result is obtained when using BOW features with TF.IDF by accuracy value 0.85. In the second place, BOW features with Binary TF have slightly diferent result with accuracy value 0.84. Meanwhile, BOW features with Raw TF comes in third place with accuracy value 0.83. Surprisingly, our proposed method has the lowest accuracy value, only 0.70.

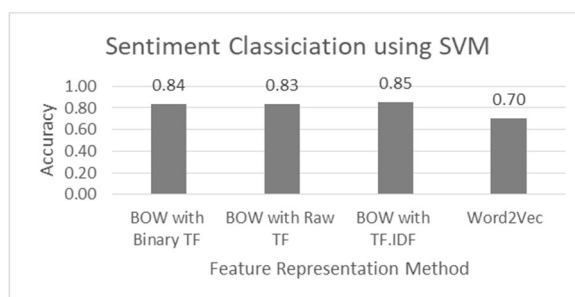


Figure 2. Experiment results

The dataset used in this experiment can be said as small dataset. In a small dataset, Word2Vec cannot capture the the semantic and syntactic information of words very well. When Word2Vec learn the word representation, each word starts at random position in the vector space. The words will be moved closer into the position of words that similar to them gradually based on their neighbors in training data. If we have very large dataset, all the words can be arranged so that all those pairwise similarities are simultaneously upheld because it have so many varied examples to gradually moved them all into better positions. Otherwise, in small dataset, there are very few examples where the words that could be similar are neighbors in training data. With very few examples where there are shared nearby-words, there's few bases for moving the all those pairwise similarities to the same position. Hence, Word2Vec cannot trained well using small dataset.

4. CONCLUSION

In this study, we used Word2Vec model to represent the features for product review sentiment classification in Indonesian language. We used SVM for the classification method. We also compared the Word2Vec based classification performance with Bag of Words features using Binary TF, Raw TF, and TF.IDF. In general, SVM can performs well on the sentiment classification. However, the Word2vec model have the lowest accuracy value than other method. This is because we only have small dataset to train the Word2Vec model. Word2Vec need large example to learn the word representation and place similar words into closer position. Otherwise, in a small dataset, there too many examples to move the words into the better place.

In the future work, we can use larger dataset to build the Word2Vec model. This dataset does not need to be labeled first as positive or negative. This dataset also does not need to be sentiment analysis dataset. We can use another dataset such as news, articles, wikipedia, and so on.

REFERENCES

- [1] Dang, Yan, Yulei Zhang, and Hsinchun Chen. "A Lexicon-Enhanced Method for Sentiment Classification: An Experiment on Online Product Reviews." *IEEE Intelligent Systems* 25, no. 4: 46-53, 2010.
- [2] Bailey, Ainsworth Anthony. "Thiscompanysucks.com: The Use of the Internet in Negative Consumer-To-Consumer Articulations." *Journal of Marketing Communications* 10, no. 3: 169-182, 2004.
- [3] Armstrong, Arthur, and John Hagel. "The Real Value of Online Communities." *Knowledge and communities* 74, no. 3 (2000): 85-95, 2000.
- [4] West, Patricia M., and Susan M. Broniarczyk. "Integrating Multiple Opinions: The Role of Aspiration Level on Consumer Response to Critic Consensus." *Journal of Consumer Research* 25, no. 1: 38-51, 1998.
- [5] Tsang, Alex SL, and Gerard Prendergast. "Is a "star" Worth a Thousand Words? The Interplay between Product-Review Texts and Rating Valences." *European Journal of Marketing* 43, no. 11/12, 1269-1280, 2009.
- [6] Hu, Nan, Indranil Bose, Noi Sian Koh, and Ling Liu. "Manipulation of Online Reviews: An Analysis of Ratings, Readability, and Sentiments." *Decision Support Systems* 52, no. 3: 674-684, 2012.
- [7] Cui, Hang, Vibhu Mittal, and Mayur Datar. "Comparative Experiments on Sentiment Classification for Online Product Reviews." In *AAAI*, vol. 6, pp. 1265-1270. 2006.
- [8] Bickart, Barbara, and Robert M. Schindler. "Internet Forums as Influential Sources of Consumer Information." *Journal of interactive marketing* 15, no. 3: 31-40, 2001.
- [9] Basuroy, Suman, Subimal Chatterjee, and S. Abraham Ravid. "How Critical Are Critical Reviews? The Box Office Effects of Film Critics, Star Power, and Budgets." *Journal of marketing* 67, no. 4: 103-117, 2003.
- [10] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up?: Sentiment Classification Using Machine Learning Techniques." In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pp. 79-86. Association for Computational Linguistics, 2002.
- [11] Liu, Bing. "Sentiment Analysis and Subjectivity." *Handbook of natural language processing* 2: 627-666, 2010.
- [12] Zhang, Dongwen, Hua Xu, Zengcai Su, and Yunfeng Xu. "Chinese Comments Sentiment Classification Based on Word2vec and SVM Perf." *Expert Systems with Applications* 42, no. 4: 1857-1863, 2015.
- [13] Fauzi, M. Ali, and Tri Afrianto. "Improving Sentiment Analysis of Short Informal Indonesian Product Reviews using Synonym Based Feature Expansion." *TELKOMNIKA (Telecommunication Computing Electronics and Control)* 16, no. 2, 2018.
- [14] Rofiqoh, Umi, Rizal Setya Perdana, and M. Ali Fauzi. "Sentiment Analysis of Indonesian Cellular Telecommunication Service Provider User Satisfaction Levels on Twitter with Supporting Vector Machine and Lexicon Based Features (in Bahasa: Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Features)." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN 2548: 964X*.
- [15] Lestari, Agnes Rossi Trisna, Rizal Setya Perdana, and M. Ali Fauzi. "Sentiment Analysis of DKI 2017 Regional Head Election Opinions on Indonesian Language Twitter Documents Using Naïve Bayes and Emoji Weighting (in Bahasa: Analisis Sentimen Tentang Opini Pilkada Dki 2017 Pada Dokumen Twitter Berbahasa Indonesia

- Menggunakan Naïve Bayes dan Pembobotan Emoji)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN* 2548: 964X.
- [16] Nurjanah, Winda Estu, Rizal Setya Perdana, and Mochammad Ali Fauzi. "Sentiment Analysis of Television Shows Based on Community Opinion on Twitter Social Media using the K-Nearest Neighbor Method and Weighting Retweet Number (in Bahasa: Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet)." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN* 2548: 964X.
- [17] Antinasari, Prananda, Rizal Setya Perdana, and M. Ali Fauzi. "Sentiment Analysis of Film Opinion on Indonesian Language Twitter Documents Using Naive Bayes with Non-Standard Word Repair (in Bahasa: Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku)" *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN* 2548: 964X.
- [18] Zhai, Zhongwu, Bing Liu, Hua Xu, and Peifa Jia. "Clustering Product Features for Opinion Mining." In *Proceedings of the fourth ACM international conference on Web search and data mining*, pp. 347-354. ACM, 2011.
- [19] Fauzi, M.A., "Random Forest Approach for Sentiment Analysis in Indonesian Language," *Indonesian Journal of Electrical Engineering and Computer Science*, 12(1), 2018.
- [20] Fauzi, M.A. and Yuniarti, A., "Ensemble Method for Indonesian Twitter Hate Speech Detection," *Indonesian Journal of Electrical Engineering and Computer Science*, 11(1), 2018.
- [21] Ghiassi, Manoochehr, James Skinner, and David Zimbra. "Twitter Brand Sentiment Analysis: A Hybrid System Using N-Gram Analysis and Dynamic Artificial Neural Network." *Expert Systems with applications* 40, no. 16: 6266-6282, 2013.
- [22] Prasanti, A.A., Fauzi, M.A. and Furqon, M.T., "Neighbor Weighted K-Nearest Neighbor for Sambat Online Classification," *Indonesian Journal of Electrical Engineering and Computer Science*, 12(1), 2018.
- [23] Taboada, Maite, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. "Lexicon-based Methods for Sentiment Analysis." *Computational linguistics* 37, no. 2: 267-307, 2011.
- [24] Davidov, Dmitry, Oren Tsur, and Ari Rappoport. "Enhanced Sentiment Learning Using Twitter Hashtags and Smileys." In *Proceedings of the 23rd international conference on computational linguistics: posters*, pp. 241-249. Association for Computational Linguistics, 2010.
- [25] Agarwal, Basant, Soujanya Poria, Namita Mittal, Alexander Gelbukh, and Amir Hussain. "Concept-level Sentiment Analysis With Dependency-Based Semantic Parsing: A Novel Approach." *Cognitive Computation* 7, no. 4: 487-499, 2015.
- [26] Lizhen, Liu, Song Wei, Wang Hanshi, Li Chuchu, and Lu Jingli. "A Novel Feature-Based Method for Sentiment Analysis of Chinese Product Reviews." *China communications* 11, no. 3: 154-164, 2014.
- [27] Singh, Vivek Kumar, Rajesh Piryani, Ashraf Uddin, and Pranav Waila. "Sentiment Analysis of Movie Reviews: A New Feature-Based Heuristic for Aspect-Level Sentiment Classification." In *Automation, computing, communication, control and compressed sensing (iMac4s), 2013 international multi-conference on*, pp. 712-717. IEEE, 2013.
- [28] Turian, Joseph, Lev Ratinov, and Yoshua Bengio. "Word Representations: A Simple and General Method for Semi-Supervised Learning." In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pp. 384-394. Association for Computational Linguistics, 2010.
- [29] Seok, Miran, Hye-Jeong Song, Chan-Young Park, Jong-Dae Kim, and Yu-seop Kim. "Named Entity Recognition Using Word Embedding As a Feature." *International Journal of Software Engineering and Its Applications* 10, no. 2: 93-104, 2016.
- [30] Lai, Siwei, Kang Liu, Shizhu He, and Jun Zhao. "How to Generate a Good Word Embedding." *IEEE Intelligent Systems* 31, no. 6: 5-14, 2016.
- [31] Mikolov, Tomas, Wen-tau Yih, and Geoffrey Zweig. "Linguistic Regularities in Continuous Space Word Representations." In *hlt-Naacl*, vol. 13, pp. 746-751. 2013.
- [32] Xue, Bai, Chen Fu, and Zhan Shaobin. "A Study on Sentiment Computing and Classification of Sina Weibo with word2vec." In *Big Data (BigData Congress), 2014 IEEE International Congress on*, pp. 358-363. IEEE, 2014.
- [33] Lilleberg, Joseph, Yun Zhu, and Yanqing Zhang. "Support Vector Machines and Word2vec for Text Classification with Semantic Features." In *Cognitive Informatics & Cognitive Computing (ICCI* CC), 2015 IEEE 14th International Conference on*, pp. 136-140. IEEE, 2015.
- [34] Joshi, Aditya, Vaibhav Tripathi, Kevin Patel, Pushpak Bhattacharyya, and Mark Carman. "Are Word Embedding-based Features Useful for Sarcasm Detection?." *arXiv preprint arXiv:1610.00883*, 2016.
- [35] Sienčnik, Scharolta Katharina. "Adapting word2vec to Named Entity Recognition." In *Proceedings of the 20th Nordic Conference of Computational Linguistics, NODALIDA 2015, May 11-13, 2015, Vilnius, Lithuania*, no. 109, pp. 239-243. Linköping University Electronic Press, 2015.
- [36] Liu, Haixia. "Sentiment Analysis of Citations Using Word2vec." *arXiv preprint arXiv:1704.00177* (2017).
- [37] Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. "Distributed Representations of Words and Phrases and Their Compositionality." In *Advances in neural information processing systems*, pp. 3111-3119. 2013.
- [38] Yang, Xiao, Craig Macdonald, and Iadh Ounis. "Using Word Embeddings In Twitter Election Classification." *arXiv preprint arXiv:1606.07006*, 2016.
- [39] Bengio, Yoshua, Réjean Ducharme, Pascal Vincent, and Christian Jauvin. "A Neural Probabilistic Language Model." *Journal of machine learning research* 3, no. Feb (2003): 1137-1155, 2003.

- [40] Fauzi, M. Ali, Agus Zainal Arifin, and Sonny Christiano Gosaria. "Indonesian News Classification Using Naïve Bayes and Two-Phase Feature Selection Model." *Indonesian Journal of Electrical Engineering and Computer Science* 8, no. 3, 2017.
- [41] Suharno, Claudio Fresta, M. Ali Fauzi, and Rizal Setya Perdana. "Indonesian Language Classification on Sambat Online Complaint Documents Using the K-Nearest Neighbors And Chi-Square Method (in Bahasa: Klasifikasi Teks Bahasa Indonesia Pada Dokumen Pengaduan Sambat Online Menggunakan Metode K-Nearest Neighbors Dan Chi-Square.)" *Systemic: Information System and Informatics Journal* 3, no. 1: 25-32, 2017.
- [42] Fauzi, M. Ali, Agus Zainal Arifin, and Anny Yuniarti. "Arabic Book Retrieval using Class and Book Index Based Term Weighting." *International Journal of Electrical and Computer Engineering (IJECE)* 7, no. 6: 3705-3710, 2017.
- [43] Fauzi, M. Ali, Djoko Cahyo Utomo, Eko Sakti Pramukantoro, and Budi Darma Setiawan. "Automatic Essay Scoring System Using N-Gram And Cosine Similarity For Gamification Based E-Learning."
- [44] Pramukantoro, Eko Sakti, and M. Ali Fauzi. "Comparative Analysis of String Similarity and Corpus-Based Similarity for Automatic Essay Scoring System on E-Learning Gamification." In *Advanced Computer Science and Information Systems (ICACSIS), 2016 International Conference on*, pp. 149-155. IEEE, 2016.
- [45] Fauzi, M. Ali, Agus Arifin, and Anny Yuniarti. "Term Weighting Based on Index of Books and Classes for Ranking of Arabic Documents (in Bahasa: Term Weighting Berbasis Indeks Buku dan Kelas untuk Perangkingan Dokumen Berbahasa Arab.)" *Lontar Komputer: Jurnal Ilmiah Teknologi Informasi* 5, no. 2, 2013.
- [46] Mikolov T, Chen K, Corrado G, Dean J. Efficient Estimation Of Word Representations In Vector Space. arXiv Preprint arXiv:1301.3781. 2013 Jan 16.
- [47] Burges, Christopher JC. "A Tutorial on Support Vector Machines for Pattern Recognition." *Data mining and knowledge discovery* 2, no. 2: 121-167, 1998.
- [48] Liu, Bing. *Web data mining: exploring hyperlinks, contents, and usage data*. Springer Science & Business Media, 2007.
- [49] McCallum, Andrew, and Kamal Nigam. "A Comparison of Event Models for Naive Bayes Text Classification." In *AAAI-98 workshop on learning for text categorization*, vol. 752, no. 1, pp. 41-48. 1998.
- [50] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research*. 2011;12(Oct):2825-30, 2011.

BIOGRAPHY OF AUTHOR



M. Ali Fauzi, male, is a lecturer in Faculty of Computer Science, Universitas Brawijaya. His research interest is Text Mining and Natural Language Processing