ILKOM Jurnal Ilmiah Vol. 15, No. 2, August 2023, pp. 353-357
Accredited 2nd by RISTEKBRIN No. 200/M/KPT/2020; E-ISSN 2548-7779 | P-ISSN 2087-1716

Research Article

Feature Space Augmentation for Negation Handling on Sentiment Analysis

Lutfi Budi Ilmawan [a,1,*]; Muladi [a,2]; Didik Dwi Prasetya [a,3]
[a] *Universitas Negeri Malang, Malang, Indonesia*
[1] *lutfi.budi.2205349@students.um.ac.id;* [2] *muladi@um.ac.id;* [3] *didikdwi@um.ac.id*
* *Corresponding author*

Article history: Received May 05, 2023; Revised June 09, 2023; Accepted June 23, 2023; Available online August 16, 2023

Abstract

One crucial issue affecting the performance of sentiment analysis tasks is negation. Handling negation involves determining the negation scope and negation cue. Feature space augmentation is one approach used to address negation. Feature space augmentation has been carried out by some previous researchers using a negation flag with the rule that the negation scope includes all words from the explicit negation cue to the punctuation mark. This study aimed to analyze the classifier's performance when negation handling was applied by adding a new rule for the negation scope. The new rule for determining the negation scope no longer took all words from the negation cue to the punctuation mark, but only considered or ignored words with certain POS tags. The results of this study showed that using the new rule for negation scope contributed to improving the performance of the classifier in sentiment analysis tasks. The proposed approach for negation handling was superior than the previous approach in terms of accuracy, precision, recall, and f1-score.

Keywords: Augmentation; Feature Space; Negation Handling; POS tag; Sentiment Analysis.

Introduction

Sentiment analysis is part of Natural Language Processing (NLP) which aims to extract sentiments and opinions from text [1], [2]. Sentiment analysis can be considered as text classification [3]–[5] because the process includes classifying either a text has positive or negative sentiments [6]. Sentiment analysis may appear to be an easy process, but it actually covers many problems in the NLP sub-task [6], one of which is negation. Negation is a linguistic phenomenon in natural language that reverses the meaning of a sentence. Usually negation will reverse the affirmative sentence to become negative, which affects the polarity of the word, so that the sentiment expressed in the text also changes [7]. Handling negation is an important sub-task in sentiment analysis in NLP and is considered as one of the most difficult problems in NLP [7], [8].

Handling negations on sentiment analysis tasks based on a machine learning approach has been carried out by previous researchers [9]–[11] by using negation tags or negation flags. The use of negation tags is included in one of the feature space augmentation approaches for negation handling [8]. Negation handling is done by adding a "NOT_" tag to every word identified as negation words (not, isn't, didn't, etc.) and the first punctuation mark after the negation word. Handling negations with negation tags was first introduced by [12]. The use of negation tags in research [10] gave quite good results, an increase in accuracy of 5.8%, a classifier accuracy value of 82.9%. Research conducted by [11] used negation tags for Chinese sentiment analysis, the results of the accuracy increased by 2.7%, the accuracy value was 85.95%.

From several studies related to negation handling using negation tags [9]–[11], the scope of the negation includes all words starting from negation cues (words of negation) until punctuation marks are found. This approach has several weaknesses, for example in the following sentence:

"This movie is not very entertaining and I hate it."

If using negation tags from research [12], the results will be like the following text, assuming that the words "and" and "is" are included in the stoplist:

"this movie NOT_very NOT_entertaining NOT_I NOT_hate NOT_it."

The placement of the “not tag” is correct in the NOT_entertaining feature, but has problems with other features, such as NOT_hate. The NOT_hate feature in this approach is interpreted as a positive sentiment, because it reverses the polarity of the sentiment of the word hate, whose initial polarity is negative. Of course the polarity of this positive sentiment does not match the original text, before giving the NOT_ tag. Based on these problems, the research would provide a new rule for the negation scope so that the NOT_ tag is in accordance with its placement, by choosing certain words based on the Part-of-Speech (POS) tag. So in this study, negation handling was carried out using the new negation scope rules. The negation tag would be applied in the sentiment analysis task where the sentiment classification process used a machine learning approach.

The machine learning method is a popular method used for sentiment analysis. According to previous studies, the accuracy of the machine learning approach is better than the lexicon based approach [13], [14]. Research [15] conducted sentiment analysis with several approaches with the aim of introducing a combination of machine learning and lexicon based approaches under the name psenti. The results of this study showed that the accuracy of the machine learning approach still had high accuracy compared to other methods, including psenti. The machine learning approach ch had an accuracy of 85.41% and the introduced pSenti method had an accuracy of 82.27%. A comparison of the results between the machine learning approach and the Semantic Orientation Approach was also carried out by [16], in terms of accuracy, the results of this study were the same as previous research [16], namely the accuracy of the machine learning approach was 91% and the accuracy of the lexicon based approach was 86%.

The purpose of this research was to measure the performance of the machine learning classifier for sentiment analysis if the negation tag with the new negation scope rule was employed.

Method

A. Preprocessing

Before the data ready to be used as training data, preprocessing was the first step that must be carried out in the sentiment analysis task, while the processes that were carried out in preprocessing sequentially as follow:

- Tweet normalization, was the stage of changing all user tags/mentions (example: @username) to 'user' and all URLs in tweets were changed to 'url'. At this stage, the case folding process was also carried out as an initial step to standardize the characters in the lower case form.
- Punctuation removal.
- Stopword elimination, deleted words contained in the stopwords list. Stopwords were considered as noise data because they had a high frequency but did not significantly affect the sentiment value of a sentence.

B. Negation Handling

The proposed negation handler was based on the study [9] with some changes. Changes were made to the negation scope. In research [9]–[11], the negation scope included all words starting from the negation cue (negation words) to punctuation, thus allowing for features or words that should not be negated, as shown in **Table 1**. In this study , the negation scope was changed with the following rule:

- If the word after the negation cue contained an adverb, then that word was omitted/ignored, then this process was repeated until the next identified word was not an adverb.
- After all adverb words had been removed, a 'NOT_' tag to the next word was added.
- Tags were only given to one feature / word.

Detecting adverbs in sentences was conducted by matching with an existing adverb list, this process did not use POS tags because it can increase computational cost in processing. The following is an example of negation handling using the proposed approach according to **Table 1**.

Tabel 1. Negation Handling Example

Sentence	Negation approach	Result
This movie is not very entertaining and I hate it.	S. R. Das and M. Y. Chen [9]	this movie NOT_very NOT_entertaining NOT_I NOT_hate NOT_it
	Proposed approach	this movie NOT_entertaining I hate it

C. Feature Model

Unigram was chosen as the used feature model. Unigram represented each word as a feature. Several studies have shown that the use of unigrams was better than other $n-gram$ models [17], [18] in supporting classifier performance.

D. Sentiment Classifier

The machine learning approach used for sentiment classification was Multinomial Naïve Bayes (MNB). The Naïve Bayes classification method was a classification method that had good performance, even in some cases compared to the deep learning approach, classification methods with traditional approaches such as SVM and naïve bayes was still able to compete well. In research [19], the naïve bayes classification method was able to outperform the deep learning approach, Multilayer Perceptron (MLP).

The determination of sentiment polarity refered to Equation 1, then a maximum value search was performed based on the final probability value of each class $P(xi|cj)$.

$$c_{map} = arg \max_{c_j \in C} P(c_j) \prod_{i=1}^{m} P(x_i|c_j) \quad (1)$$

In Equation 2, many conditional probabilities were multiplied, this could produce fractional values that can exceed the limit of the existing floating point value [20].

$$c_{map} = arg \max_{c_j \in C} \log P(c_j) + \sum_{i=1}^{m} P(x_i|c_j) \quad (2)$$

Therefore, it was better to perform calculations by adding the logarithm of the probabilities instead of multiplying the probabilities. The class with the highest probability log score still showed the most probable outcome.

Results and Discussion

A. Data

All datasets used to evaluate classifier performance came from research [21]. The dataset for the training process was 1,600,000 tweets and the dataset for testing was 359 tweets which were manually labeled.

B. Measuring classifier performance

The test scenario was carried out by comparing the approach for proposed negation handling adopted from research [12] and the results of research conducted by [21] which overcame negations by using the bigram feature model. Overall the results of classifier performance measurements can be seen in **Table 2**.

Tabel 2. Comparison of Classifier Accuracy

Negation Approach	Feature Model	Accuracy Result of Naïve Bayes Classifier
Proposed approach	Unigram	**83%**
S. R. Das and M. Y. Chen [9]	Unigram	79%
A. Go, R. Bhayani, and L. Huang [21]	Unigram + Bigram	82.7%

The results of performance measurements: precision, recall, and f1-score can be seen in **Table 3**. The table did not include results from research [21] because in that research, classifier performance was only tested using accuracy testing.

Tabel 3. Evaluation result of precision, recall, dan f1-score

Negation Approach	Class	Precision	Recall	F1-Score
Proposed approach	Negative	83%	83%	83%
	Positive	83%	84%	83%
S. R. Das and M. Y. Chen [9]	Negative	80%	77%	79%
	Positive	78%	82%	80%

The classifier performance test results showed that the proposed negation tag approach was superior than the original negation tag approach by [12]. Likewise, when compared with research [21] which, in addition to using unigrams for its feature models, this study also used bigrams, but the performance of the proposed approach was still better.

Future research can use word representation to produce more complex features such as TF-IDF and word embedding, because in this study the word representation only used a simple approach, namely Bag-of-Words. Further development is also expected for the application of the feature selection algorithm at one of the stages in the feature extraction process such as information gain, chi-squared, or mutual information, because the use of feature selection in several previous studies can improve classifier performance [22]–[25].

Conclusion

The application of negation tags for negation handling by ignoring adverbs can improve classifier performance on sentiment analysis tasks. Based on the test results, there was an increase in classifier performance in terms of accuracy, precision, recall, and f1-score when negation handling was carried out using the proposed approach. An accuracy result of 83% was achieved using the proposed approach, which was better than previous studies, as well as the test results in terms of precision, recall, and F1-score.

References

[1] I. Chaturvedi, E. Cambria, R. E. Welsch, and F. Herrera, "Distinguishing between facts and opinions for sentiment analysis: Survey and challenges," *Inf. Fusion*, vol. 44, pp. 65–77, Nov. 2018, doi: 10.1016/J.INFFUS.2017.12.006.

[2] B. Liu, "Sentiment analysis: Mining opinions, sentiments, and emotions," *Sentim. Anal. Min. Opin. Sentim. Emot.*, pp. 1–367, Jan. 2015, doi: 10.1017/CBO9781139084789.

[3] B. Schuller, A. E. D. Mousa, and V. Vryniotis, "Sentiment analysis and opinion mining: on optimal parameters and performances," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 5, no. 5, pp. 255–263, Sep. 2015, doi: 10.1002/WIDM.1159.

[4] S. Ju and S. Li, "Active learning on sentiment classification by selecting both words and documents," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7717 LNAI, pp. 49–57, 2013, doi: 10.1007/978-3-642-36337-5_6/COVER.

[5] Y. Choi and H. Lee, "Data properties and the performance of sentiment classification for electronic commerce applications," *Inf. Syst. Front.*, vol. 19, no. 5, pp. 993–1012, Oct. 2017, doi: 10.1007/S10796-017-9741-7/FIGURES/7.

[6] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowledge-Based Syst.*, vol. 226, p. 107134, 2021, doi: 10.1016/j.knosys.2021.107134.

[7] P. K. Singh and S. Paul, "Deep Learning Approach for Negation Handling in Sentiment Analysis," *IEEE Access*, vol. 9, pp. 102579–102592, 2021, doi: 10.1109/ACCESS.2021.3095412.

[8] J. Barnes, E. Velldal, and L. Ovrelid, "Improving sentiment analysis with multi-task learning of negation," *Nat. Lang. Eng.*, vol. 27, no. 2, pp. 249–269, 2021, doi: 10.1017/S1351324920000510.

[9] S. R. Das and M. Y. Chen, "Yahoo! for amazon: Sentiment extraction from small talk on the Web," *Manage. Sci.*, vol. 53, no. 9, pp. 1375–1388, 2007, doi: 10.1287/mnsc.1070.0704.

[10] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques," no. July, pp. 79–86, 2002, [Online]. Available: http://arxiv.org/abs/cs/0205070

[11] H. Y. Lee and H. Renganathan, "Chinese Sentiment Analysis Using Maximum Entropy," in *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology*, 2011, pp. 89–93. [Online]. Available: http://acl.eldoc.ub.rug.nl/mirror/W/W11/W11-37.pdf#page=105

[12] S. Das and M. Chen, "Yahoo! for Amazon: Extracting market sentiment from stock message boards," in *Proceedings of the Asia Pacific finance association annual conference (APFA)*, 2001, vol. 35, p. 43.

[13] P. D. Turney, "Thumbs up or thumbs down? Semantic Orientation applied to Unsupervised Classification of Reviews," in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2002, no. July, pp. 417–424. doi: 10.3115/1073083.1073153.

[14] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-Based Methods for Sentiment Analysis," *Comput. Linguist.*, vol. 37, no. 2, pp. 267–307, 2011, doi: 10.1162/COLI_a_00049.

[15] A. Mudinas, D. Zhang, and M. Levene, "Combining lexicon and learning based approaches for concept-level sentiment analysis," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2012. doi: 10.1145/2346676.2346681.

[16] M. D. Devika, C. Sunitha, and A. Ganesh, "Sentiment Analysis: A Comparative Study on Different Approaches," *Procedia Comput. Sci.*, vol. 87, no. December, pp. 44–49, 2016, doi:

10.1016/j.procs.2016.05.124.

[17] A. Nasser, K. Dinçer, and H. Sever, “Investigation of the Feature Selection Problem for Sentiment Analysis in Arabic Language,” *Res. Comput. Sci.*, vol. 110, no. 1, pp. 41–54, 2016, doi: 10.13053/rcs-110-1-4.

[18] P. H. Shahana and B. Omman, “Evaluation of features on sentimental analysis,” *Procedia Comput. Sci.*, vol. 46, no. Icict 2014, pp. 1585–1592, 2015, doi: 10.1016/j.procs.2015.02.088.

[19] C. Diamantini, A. Mircoli, D. Potena, and E. Storti, “Social information discovery enhanced by sentiment analysis techniques,” *Futur. Gener. Comput. Syst.*, vol. 95, pp. 816–828, 2019, doi: 10.1016/j.future.2018.01.051.

[20] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Modern Information Retrieval (2nd edition)*, Online Edi. Cambridge: Cambridge University Press, 2009. doi: 10.1108/00242530410565256.

[21] A. Go, R. Bhayani, and L. Huang, “Twitter Sentiment Classification using Distant Supervision,” California, 2009. doi: 10.1016/j.sedgeo.2006.07.004.

[22] L. P. Hung, R. Alfred, and M. H. A. Hijazi, “A review on feature selection methods for sentiment analysis,” *Adv. Sci. Lett.*, vol. 21, no. 10, pp. 2952–2956, 2015, doi: 10.1166/asl.2015.6475.

[23] D. S. Guru, M. Ali, and M. Suhil, “A novel feature selection technique for text classification,” *Adv. Intell. Syst. Comput.*, vol. 813, pp. 721–733, 2019, doi: 10.1007/978-981-13-1498-8_63.

[24] H. M. K. Kumar and B. S. Harish, “A New Feature Selection Method for Sentiment Analysis in Short Text,” *J. Intell. Syst.*, vol. 29, no. 1, pp. 1122–1134, 2020, doi: 10.1515/jisys-2018-0171.

[25] M. Iqbal, M. M. Abid, M. N. Khalid, and A. Manzoor, “Review of feature selection methods for text classification,” *Int. J. Adv. Comput. Res.*, vol. 10, no. 49, pp. 138–152, 2020, doi: 10.19101/ijacr.2020.1048037.