Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 7220–7225
Marseille, 20-25 June 2022

Using Convolution Neural Network with BERT for Stance Detection in Vietnamese

Oanh Thi Tran*, Anh Cong Phung, Bach Xuan Ngo****
*International School, Vietnam National University, Hanoi

**Posts and Telecommunications Institute of Technology, Vietnam
oanhtt@isvnu.vn, anhpc@ptit.edu.vn, bachnx@ptit.edu.vn

Abstract

Stance detection is the task of automatically eliciting stance information towards a specific claim made by a primary author. While most studies have been done for high-resource languages, this work is dedicated to a low-resource language, namely Vietnamese. In this paper, we propose an architecture using transformers to detect stances in Vietnamese claims. This architecture exploits BERT to extract contextual word embeddings instead of using traditional word2vec models. Then, these embeddings are fed into CNN networks to extract local features to train the stance detection model. We performed extensive comparison experiments to show the effectiveness of the proposed method on a public dataset. Experimental results show that this proposed model outperforms the previous methods by a large margin. It yielded an accuracy score of 75.57% averaged on four labels. This sets a new SOTA result for future research on this interesting problem in Vietnamese.

Keywords: stance detection, Vietnamese, BERT, CNN

1. Introduction

With the development of information technology, people nowadays can easily express their stances against specific topics on social media like sites, news portals, forums, and online newspapers. Stance (Dilek and Fazli, 2020) is defined as the expression of the speaker's attitude, standpoint, and judgement toward a target. The analysis of these comments is typically used as an initial step for fake news detection (Sun et al., 2018) or in understanding public reaction towards the social issues mentioned in these claims. Stance detection is a newly emerging research field with the goal is to identify the stance of the text authors toward a target which is either explicitly mentioned or implied within the text. An example is given in Figure 1, which shows a headline of a news article and its four comments. This article received lots of the follow-up comments which indicate the reaction of public users (i.e. *agree*, *disagree*, *discuss* or *unrelated*) towards the claim mentioned.

To deal the task, researchers framed it as a classification problem, and then exploit different machine learning (ML) methods to solve it. So far, there existed many works proposed to solve the task using features-based ML approaches (Kucher et al., 2018; Simaki et al., 2017; Swami et al., 2017). At present, deep learning approaches (Chung et al., 2014; Sun et al., 2018; Zhang et al., 2017) have been extensively investigated and yielded better performance for this task.

There were also many datasets in popular languages made publicly available for research community via official competitions or challenges like *SemEval-2016 Task 6: Detecting Stance in Tweets*[1], *Shared Task of Stance Detection in Chines Microblogs at NLPCC-ICCPOL-2016*[2], *Shared task of stance detection in Spanish and Catalan tweets at IberEval-2017*[3], and *Fake News Challenge Stage 1: Stance Detection*[4]. These have boosted research on stance detection by providing annotated datasets, evaluation metrics, and a wide range of proposed methods.

While most current works have been conducted for high-resource languages like English, Catalan, and Chinese, we have seen very little attention to date for low-resource languages. To narrow down this gap, this work is dedicated to Vietnamese stance detection to develope a robust stance detection model for Vietnamese. We introduce a model using BERT (Devlin et al., 2019) to encode token embeddings, then feeding these embeddings into CNN networks (LeCun and Bengio, 1998) to learn a sequence modelling task. Multiple types of comparison experiments are also set up and compared with the proposed model. According to the final experimental results, the proposed model is very robust in comparison to other models. It outperforms other previous methods by a large margin. It yielded the accuracy score of 75.57% and provided a strong baseline for future research.

Our paper makes the following contributions:

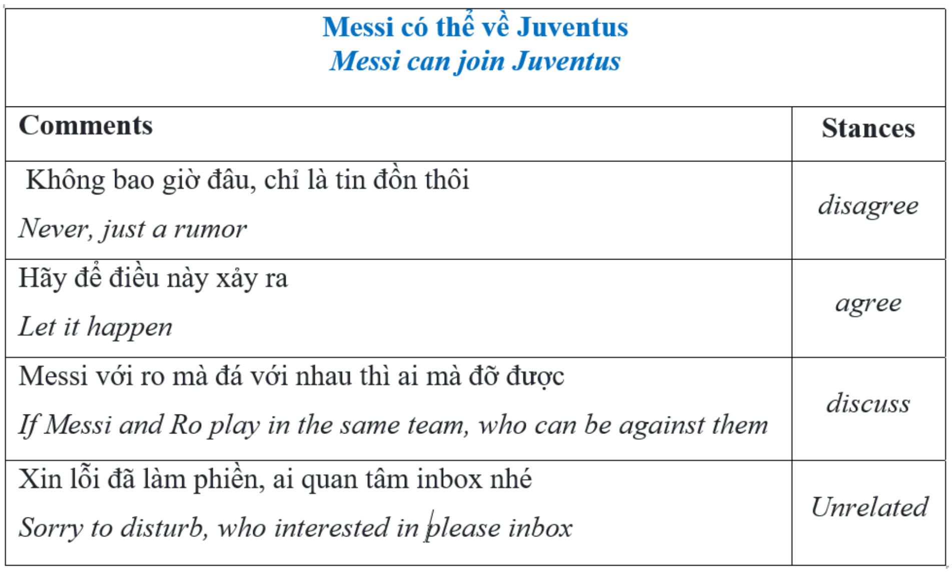
- Present a systematic study on stance detection for a low-resource language.
- Perform extensive experiments and report the SOTA result for future research on this interesting direction.

[1] https://alt.qcri.org/semeval2016/task6/

[2] http://tcci.ccf.org.cn/conference/2016/

[3] https://stel.ub.edu/Stance-IberEval2017/

[4] http://www.fakenewschallenge.org/



Messi có thể về Juventus
Messi can join Juventus

Comments	**Stances**
Không bao giờ đâu, chỉ là tin đồn thôi *Never, just a rumor*	*disagree*
Hãy để điều này xảy ra *Let it happen*	*agree*
Messi với ro mà đá với nhau thì ai mà đỡ được *If Messi and Ro play in the same team, who can be against them*	*discuss*
Xin lỗi đã làm phiền, ai quan tâm inbox nhé *Sorry to disturb, who interested in please inbox*	*Unrelated*

Figure 1: A Vietnamese example about one claim with four comments and their corresponding stances towards this claim (*the English translations* are given right below the Vietnamese sentences and in *italics*).

The remainder of this paper is organized as follows: Related work is described in Section 2. Section 3 formally states the problem and presents the proposed methods. Section 4 first shows the dataset. Then, experimental setups, network training and experimental results are reported. Some error analysis and discussions are also mentioned in this section. Finally, we conclude the paper and point out some future work in Section 5.

2. Related Work

Approaches to stance detection can be classified into three main types (Dilek and Fazli, 2020): (1) feature-based machine learning approaches, (2) deep learning approaches, and (3) ensemble learning approaches.

Using the first machine learning approach, some typical methods are exploited both in earlier work as well as in recent stance detection competitions such as SVMs (Swami et al., 2017), Logistic regression (Kucher et al., 2018), decision tree (Simaki et al., 2017), etc. These methods require a good feature set to train the robust models. This feature set is mostly designed manually such as lexical features, word vector representation, topic modelling related features such as LDA, LSA, or TF-IDF, features based on POS tag, named entities, dependency information, coreference resolution, etc.

The disadvantage of the first approach is that building features by hands is time and cost consuming. Therefore, researchers proposed the second approach based on deep learning techniques to overcome this weakness. This approach usually did not require feature engineering by hands. For example, types of RNNs such as LSTMs (Sun et al., 2018), GRUs (Chung et al., 2014), CNNs (Zhang et al., 2017) are extensively exploited in different research to detect stances.

The third approach is ensemble learning aiming at combining the strength of multiple individual classifiers in one. For example, Tsakalidi et al. (Tsakalidis et al., 2018) used random forest is an ensemble learning algorithm that combines several decision trees to cover the training dataset. Zhou et al. (Zhou et al., 18–32) proposed a combination of bidirectional GRU and CNN with an attention mechanism. It was reported to outperform the SVM baseline (and the best performing approach) of SemEval-2016 shared task. Zhang et al. (Zhang et al., 2017) combined LSTMs and CNNs for detecting stances in debates.

As can be seen that most studies on stance detection have been performed on higher resource languages (Lai et al., 2020) like English, Chinese, French, Italian, Catalan and Spanish, etc. This is due to the availability of many annotated datasets and many public competitions such as:

- *English*: the dataset of tweets created within the course of the SemEval-2016 task.
- *Chinese*: The dataset of microblogs for NLPCC-ICCPOL-2016 stance detection task.
- *Catalan* and *Spanish*: The dataset of tweets compiled for IberEval-2017 stance detection task.

This paper is dedicated to the task of analyzing stances in the Vietnamese social issues which cover a wide range of social topics. Therefore, this work will stimulate the follow-up research on this interesting yet unexplored problem in Vietnamese.

3. A Proposed Architecture using Transformer with CNN networks

This model utilizes knowledge embedded in pre-trained BERT language models by feeding the contextualized embeddings of the last hidden layers into a several filters and convolution layers of the CNN. CNN is an ideal replacement for LSTM/GRUs (Hochreiter and Schmidhuber, 1997) for sequence modelling tasks. This network can help to automatically learn semantic representation of the word sequences. It has shown a lot of achievement for many text classification tasks in popular languages. Therefore, we also hypothesis that the similar success can be gained for the Vietnamese language. Figure 2 shows the architecture to solve the task that includes the following layers:

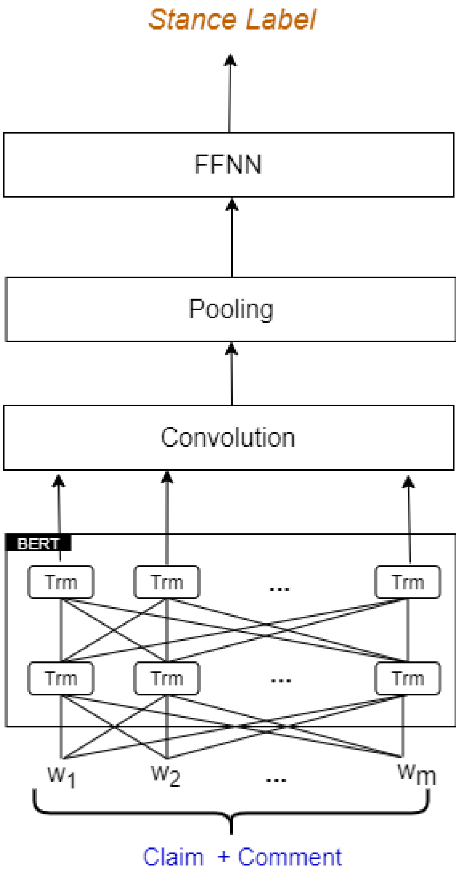


Figure 2: A framework for Vietnamese stance detection using CNN with BERT.

- **Embedding layer**: BERT is deeply bidirectional, unsupervised language representation, pre-trained using only a plain text corpus. This contextual model generates a representation of each subword based on the other sub-words in the sentence. Sub-words in text are separated using Byte-Pair Encoding subword algorithm. We extract the weights from the last hidden layer of the BERT model (Devlin et al., 2019) and use them as embeddings to train the stance detection task using CNNs (LeCun and Bengio, 1998).

- **Texts representation using CNNs**: This model exploits CNNs to model the interaction between different words in the claims. CNNs allow the model to extract local information between a target word and its neighbors, and hence leverages the local contexts based on n-gram word embeddings via CNNs. We extract each word's local features with several kernel sizes. The kernel size is equal to the size of a convolutional window across k tokens. The input of CNNs is the word vector obtained by the BERT pre-trained language model.

- **Feed forward neural network**: The output of CNNs is passed on to the next fully-connected neural network layer to yield the final context predictions for the whole texts.

Table 1: Number of comments pertaining to each stance.

No.	Stances	Comments
1	Agree	2,941
2	Disagree	2,574
3	Discuss	3,334
4	Unrelated	2,404
Total: 11,253		

4. Experiments

4.1. Corpus

In this section, we briefly introduce the details about the annotation process for building the corpus published by (Tran et al., 2021).

They underwent five key steps to build the corpus. These data was collected from online regular newspaper and the official fan-pages of regular newspapers on Facebook. Each claim-comment pair was labeled based on whether the comment c under claim p agreed, disagreed with p made by the primary user, just gave further discussion about p, or have no relation to p.

Two annotators carried out the annotation. After the re-checking process, they also calculated the Cohen's Kappa coefficient to measure the inter agreement. The value was 89.2% which indicates almost perfect agreement between them. Some statistics about the data is given in Figure 1. The number of claims is 500.

4.2. Data preprocessing

Vietnamese social media texts pose new challenges because they are usually short, informal, full of misspellings, abbreviation, slang words, etc. Hence, before building the model, a set of text processing steps is necessary as follows:

- Remove special characters (e.g = , ¡ , @ , $, :)), and icons.

- Split sticky words (e.g '*toi dongy* / (I agree)' is split to '*toi dong y*').

- Correct elongate words (e.g '*aloоооо*' (*hello*) is replaced by '*alo*').

- Replace typical words with their correct ones (e.g one negation word, '*không/no*', can be written as '*khong*', '*ko*', '*khg*', '*k*', etc).
- Vietnamese words are not split by white spaces. For this reason, we also performed word segmentation using the Pyvi library[5].

4.3. Experimental setups

We divided the corpus into train-dev-test sets with the rates 7:1:2 and reported the most commonly used evaluation metrics such as precision, recall, F_1, and accuracy scores. Word vectors were initialized using word2vec word embeddings with 300 dimensions. The size of hidden units in LSTM, GRU was set to 200. The dropout rate was 0.2 for LSTM, CNN layers. The learning rate of Adam optimizer was 0.001. The batch size was optimized in the range of (16, 32, 64, and 128). For CNN, we used different kernel sizes of 2,3,4 and 5. For BERT, we exploited PhoBERT[6] optimized for Vietnamese based on Roberta.

4.4. Experimental Results

In this section, we report the accuracy score on all labels and the precision, recall, and F_1 scores on each label using the following experimental settings:

- *word2vec-GRU, word2vec-LSTM, word2vec-CNN*: The input text is represented by the *word2vec*[7] model, and these word vectors are fed to the GRU, LSTM and CNN models for feature extraction and classification.
- *BERT-finetuned*: The corresponding word vectors are trained by BERT model for the input text, then finetuned on the stance detection data.
- *BERT-LSTM, BERT-GRU, BERT-CNN*: The corresponding word vectors were trained by BERT model, which were then classified by LSTM, GRU or CNN neural networks.

4.4.1. Effectiveness of different deep learning methods

Table 2 shows experimental results of the investigated methods.

The experimental results of using BERT-based and word2vec-based models suggested that using BERT is more effective than using word2vec in representing words with contextual information. The best *BERT-CNN* model remarkably boosted the accuracy score from 64.57% of the *word2vec-GRU* to 75.57%. Therefore it should be preferable to extract word embeddings using BERT for this task.

Table 2 (lower part) showed experimental results of four models based on BERT. It can be seen that among these four models, CNN model had a better effect in obtaining the semantic local features of text compared with LSTM, GRU networks and using BERT on its own with fine-tuning techniques. We saw a considerable improvement over BERT on its own (by 1.75%), BERT-GRU (by 1.52%) and BERT-LSTM (by 0.98%). Overall, the best model, *BERT-CNN*, significantly outperformed other methods by a large margin and yielded 75.57% in the accuracy score.

Table 2: Experimental results of the proposed methods averaged on four stance labels.

Number	Methods	Accuracy
Best previous work(Tran et al., 2021)		
0	biLSTM+Att+rich-features	66.32
word2vec embeddings		
1	word2vec-GRU	64.57
2	word2vec-LSTM	67.24
3	word2vec-CNN	69.79
BERT embeddings		
4	BERT-fine-tuning	73.82
5	BERT-GRU	74.05
6	BERT-LSTM	74.59
7	BERT-CNN	**75.57**

Table 3: Experimental results of the best *BERT-CNN* model on four stance labels.

Stance Labels	Precision	Recall	F_1 score
Agree	76.77	79.67	78.19
Disagree	72.73	59.74	65.62
Discuss	64.87	72.71	68.57
Unrelated	91.62	89.37	90.48

Looking at the first row in Table 2 showing the experimental results of the best previous model using attentive biLSTM with rich features sets(Tran et al., 2021), we acknowledge a remarkable improvement on the accuracy score. The new proposed model enhanced the accuracy score by a large margin of more than 9%.

4.4.2. Experimental results of the best method on four stance labels

Table 3 shows the experimental results of the best neural model, *BERT-CNN*, on four stance labels. We can see that the *Unrelated* label is the easiest label to make prediction with the F_1 score of 90.48%. The reason is that most comments of this type didn't share many common keywords with the claim. Among three remaining stance labels, the *disagree* and *discuss* labels are the most difficult to predict. The comments belonging to these stances were usually longer than other stances' by providing more evidence to strengthen the users' opinions.

4.5. Error Analysis and discussion

From the predicted labels of the model *BERT-CNN*, we performed analyzing typical errors generated. Table 4

[5] https://github.com/trungtv/pyvi

[6] https://github.com/VinAIResearch/PhoBERT

[7] https://github.com/sonvx/word2vecVN

Gold stance	wrongly-predicted stances		
discuss	agree 30.56	disagree 56.42	unrelated 12.89
agree	disagree 25.43	discuss 53.11	unrelated 21.42
disagree	agree 30.11	discuss 51.36	unrelated 18.4
unrelated	agree 35.16	disagree 15.89	discuss 48.91

Table 4: The most wrongly-predicted stance labels by the best stance detection model.

shows the statistics about these errors. The second column of the table presents the wrong stances predicted for the given gold stance observed. We acknowledged several typical error types as follows:

- The most popular errors were incorrectly assigning all the remaining stance labels into the *discuss* label with more than 50% of times.
- The stance *discuss* was wrongly predicted into *disagree* with more than 50% of times, followed by *agree* (with 30.56%), and *unrelated* (with only 12.89%).
- There is the lowest chance of wrongly predicted into *unrelated* from other gold stance labels.
- Otherwise, the other stance labels also have an average chance of mis-recognizing into *agree* or *disagree* labels.

5. Conclusion

We have witnessed a lot of studies dedicated to high-resource languages due to the availability of annotated data. Unfortunately, for low-resource languages, namely Vietnamese, to the best of our knowledge, this problem has received little attention so far. There is only one published work on this problem due to the lack of available dataset. Hence, to further stimulate this research, this paper focused on proposing a robust and effective model to detect stances in Vietnamese. Based on the published corpus(Tran et al., 2021), we conducted extensive experiments to make comparison to the previous models. The experimental results showed that the model exploiting BERT with CNN outperformed other strong baseline methods by a large margin. It yielded 75.57% accuracy score on the test set. This is not an easy task so this result would provide a new challenge for future research on this interesting problem in Vietnamese.

In the future, lots of work should be done to improve the performance. For example, more features should be investigated to enrich the model. Moreover, we can also exploit other architectures in trying to enhance the performance.

6. Acknowledgements

7. Bibliographical References

Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. In *arXiv preprint arXiv:1412.3555*.

Devlin, J., Chang, M., Lee, K., and Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL, Volume 1*, pages 4171–4186.

Dilek, K. and Fazli, C. (2020). Stance detection: A survey. *ACM Comput. Surv. Journal*, aricle number 12:37 pages.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Journal Neural Computation*, 9 (8):1735–1780.

Kucher, K., Paradis, C., and Kerren, A. (2018). Visual analysis of sentiment and stance in social media texts. In *20th EG/VGTC Conference on Visualization*.

Lai, M., Cignarella, A. T., Farías, D. I. H., Bosco, C., Patti, V., and Rosso, P. (2020). Multilingual stance detection in social media political debates. *Computer Speech and Language*.

LeCun, Y. and Bengio, Y., (1998). *Convolutional networks for images, speech, and timeseries*, pages 255–258. MIT Press Cambridge, USA.

Simaki, V., Paradis, C., and A.Kerren. (2017). Stance classification in texts from blogs on the 2016 british referendum. In *Int Conf. on Speech and Computer*, pages 700–709.

Sun, Q., Wang, Z., Zhu, Q., and Zhou, G. (2018). Stance detection with hierarchical attention network. In *Int Conf. on Computational Linguistics*, pages 2399—-2409.

Swami, S., Khandelwal, A., Shrivastava, M., and Sarfaraz-Akhtar, S. (2017). Ltrc-iiith at ibereval 2017: Stance and gender detection in tweets on catalan independence. In *the 2nd Workshop on Evaluation of HLT for Iberian Languages*.

Tran, O., Dao, T., and Dang, Y. (2021). Stance detection on vietnamese social media. In *The International Conference on Soft Computing and Pattern Recognition (SocPar)*.

Tsakalidis, A., Aletras, N., Cristea, A. I., and Liakata, M. (2018). Nowcasting the stance of social media users in a sudden vote: The case of the greek referendum. In *the ACM Int. Conf. on Information and Knowledge Management*.

Zhang, S., Qiu, L., Chen, F., Zhang, W., Yu, Y., and Elhadad, N. (2017). We make choices we think are going to save us: Debate and stance identification for online breast cancer cam discussions. In *Int Conf. on WWW Companion*, pages 1073—-1081.

Zhou, Y., Cristea, A. I., and Shi, L. (18—32). Connecting targets to tweets: Semantic attention-based model for target-specific stance detection. In *Int*

Conf. on Web Information Systems Engineering, page 2017.