

Improving Vietnamese Fake News Detection based on Contextual Language Model and Handcrafted Features

Khoa Pham Dang^{1,2}, Dang Van Thin^{1,2}, Ngan Luu Thuy Nguyen^{1,2,*}



Use your smartphone to scan this QR code and download this article

ABSTRACT

Introduction: In recent years, the rise of social networks in Vietnam has resulted in an abundance of information. However, it has also made it easier for people to spread fake news, which has done a great disservice to society. It is therefore crucial to verify the reliability of news. This paper presents a hybrid approach that uses a pretrained language model called vELECTRA along with handcrafted features to identify reliable information on Vietnamese social network sites. **Methods:** The present study employed two primary approaches, namely: 1) fine-tuning the model by utilizing solely textual data, and 2) combining additional meta-data with the text to create an input representation for the model. **Results:** Our approach performs slightly better than other refined BERT methods and achieves state-of-the-art results on the ReINTEL dataset published by VLSP in 2020. Our method achieved a 0.9575 AUC score, and we used transfer learning and deep learning approaches to detect fake news in the Vietnamese language using meta features. **Conclusion:** With regards to the results and analysis, it can be inferred that the number of reactions a post receives, and the timing of the event described in the post are indicative of the news' credibility. Furthermore, it was discovered that BERT can encode numerical values that have been converted into text.

Key words: fake news detection, transfer learning, pretrained language model, deep learning

INTRODUCTION

In today's world, the internet is rife with an overwhelming amount of information, and social networking sites such as Facebook and Twitter are experiencing a significant surge in popularity and becoming more accessible to the general public. However, some individuals exploit these platforms by spreading unreliable information for their personal gain, such as earning more revenue from website clicks or influencing the community's political agenda¹.

Due to the sensational nature of fake news, it is often difficult to detect its provocative intention, and if social media users do not read it carefully, fake news can spread rapidly, causing dire consequences for its victims, such as tarnishing people's reputations. Major global events are closely associated with fake news, and since the outbreak of the COVID-19 pandemic in 2020, fake news related to the pandemic has been rampant. People living in countries with travel restrictions amid COVID-19 tend to believe rumors circulated within their communities without considering their validity. In their fear of the outbreak, readers tend to trust unconfirmed information about pandemic prevention on social media, which can be reckless². This misinformation can lead to terrifying misunderstandings about any global event. Therefore, de-

termining the reliability of news has gained remarkable attention to prevent the wave of false information.

In this study, we focused on classifying trustworthy news written in Vietnamese, which is a low-resource language. Therefore, we attempted to find the best solution to predict whether the news is reliable or unreliable by implementing Transformer-based architectures and conducting several experiments on fine-tuning strategies to improve the model's performance.

RELATED WORKS

Fake news detection can be seen as a text classification task in which the input is a news article labeled either fake or real. Various machine learning and deep learning approaches have been attempted to detect fake news. Agarwal *et al.* (2021)³ proposed adding the Bi-LSTM layer with attention to the contextual embedding layers of sentences for classifying trustworthy news in the English language. Monti *et al.* (2019) studied the efficiency of graph neural networks⁴. Their architecture used a hybrid user activity and article's information representations to feed into a four-layer Graph CNN network for predicting fake or real news.

¹University of Information Technology, Ho Chi Minh city, Vietnam

²Vietnam National University, Ho Chi Minh city, Vietnam

Correspondence

Ngan Luu Thuy Nguyen, University of Information Technology, Ho Chi Minh city, Vietnam

Vietnam National University, Ho Chi Minh city, Vietnam

Email: 18520930@gm.uit.edu.vn

History

- Received: 2022-04-20
- Accepted: 2023-03-13
- Published: 2023-6-30

DOI :

<https://doi.org/10.32508/stdj.v26i1.3927>



Copyright

© VNUHCM Press. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.



Cite this article : Dang K P, Thin D V, Nguyen N L T. **Improving Vietnamese Fake News Detection based on Contextual Language Model and Handcrafted Features.** *Sci. Tech. Dev. J.* 2023; 26(2):2705-2712.

Qi *et al.*, 2019 emphasized the vital role of visual content and proposed a deep learning approach called the multidomain visual neural network to differentiate between real and fake news based on a single image⁵. They designed a CNN-based network to capture patterns of fake news images in the frequency domain and used a joint CNN-RNN model to extract image features from different semantic levels in the pixel domain. An attention mechanism was employed to fuse the feature representations of frequency and pixel domains to detect fake news in Chinese.

In the Vietnamese domain, several works have been performed by Thanh and Van (2020), Trinh *et al.* (2021), Tran *et al.* (2020)⁶⁻⁸ and Nguyen *et al.* (2020)⁹ in a VLSP shared task. They leveraged the strength of Transformer-based models and fine-tuned them on linguistic features. Additionally, Hieu *et al.* (2020)¹⁰ applied PhoBERT embeddings to extract document content and used TF-IDF to encode them into vector representations fused with vectors of meta-data. These representations were used as inputs into tree-based methods, which are Random Forest, CatBoost and LightGBM, to predict label probabilities.

METHODOLOGY

In this section, we will present our approach to solve this problem. We chose the SOTA techniques, including the BERT model and its variants that we mentioned below. We will promote a new approach that describes meta-data provided in datasets as verbal words to enrich text information. As we noticed that numeric features can contribute a lot insights to our model and BERT can somewhat encode numbers well in medium ranges. Therefore, we will describe their values as text and combine it with the original text. Ultimately, we will put the combined text into our model and using advanced fine-tuning techniques for the text input.

Data processing

Fake news detection can encompass not only the contents of news but also related information. In the ReINTEL dataset, we applied preprocessing techniques to ensure the cleanliness of the data.

- **Extract timestamp features:** We extracted timestamp features into days, months, years, hours and weekdays to enrich features for the dataset. These features were found to be helpful in the prediction accuracy of the models through data analysis.

- **Remove absurd data points:** During the data collection process, there might be several outliers. As described in Table 1, the maximum value is greater than one billion, which is implausible.
- **Fill missing data:** After removing implausible data points, a significant number of missing values were encountered, as shown in Table 2. We decided to fill missing values with different strategies: mean and median value for three features involved in interaction including number of likes, number of comments and number of shares. Furthermore, we duplicated the feature values of the DateTime data for the most similar content compared to the text, including the missing DateTime value.

For the post message, we must pre-process it more carefully than the rest of the features because it is the most crucial feature for the model to understand the text's content. The more thoroughly we process text, the more meaningful insights our models can extract and therefore improve their performance. Thus, we performed these preprocessing techniques on text content before loading them into the model:

- HTML tags, redundant characters and stop words were removed.
- Converted synonyms indicating one object into one consistent term. For instance, "Hoa Kỳ" was changed to "Mỹ", meaning "USA" in Vietnamese, and "COVID-19" and "coronavirus" were changed to "COVID".
- PhoBERT¹¹ requires word segmentation, so we use VnCoreNLP¹² to segment the input.
- We converted emojis and icons into the respective text describing them. For example, "☹️(" into the token "negative".

Data Analysis

To detect fake news, the content of news is a crucial element in determining its veracity. Therefore, we visualized the news content using word clouds to analyze and compare the differences between real and fake news. Based on Figure 1, we identified keywords related to the COVID-19 pandemic and noted some distinctions between reliable and unreliable news. Specifically, we observed the following:

- Real news content tends to contain many URLs and hashtags, indicating a greater number of reliable and accredited sources.

Table 1: Meta-data description in the experimental dataset.

	Num like post	Num comment post	Num share post
Count	4372	4372	4372
mean	1456400000000	1088786	362830.6
Std	48063100	41548540	23951710
Min	1	0	0
25%	9	0	2
50%	38	43	22
75%	432	33	87
max	1592811000	1591690000	1583711400

Table 2: Statistics of missing values in the dataset

Feature	Train set	Test set
Id	0	0
user_name	0	0
post_message	1	0
Num_like_post	115	616
Num_comment_post	10	677
Num_share_post	725	742
image	3085	1138

- Fake news content contains fewer credible sources but more keywords and negative icons. Additionally, since fake news is often fraught with a political purpose, we observed that the words ‘USA’ and ‘China’ appeared frequently in the word cloud. These two countries are global superpowers with opposing interests, making them the subject of many baseless speculations.

In addition to news content, we would like to plot the distribution of the number of reactions by fake and real news in Figure 2 after many preprocessing steps for these handcrafted features.

As we said in the previous section, fake news often has dramatically sensational content, which attracts many people to comment and lets them believe to share because readers have been overwhelmed emotionally without analyzing the content. For number of likes, those readers will read carefully until they agree with the content before deciding to press the “Like” button on that post. Following Qazvinian *et al.* (2011)⁵ and Jin *et al.* (2016)¹³, they also found that common readers tend to express their biases and feelings toward fake news more than reliable news. Furthermore, we see that unreliable news often appears from

6 PM to 10 PM, the period when people mostly surf the internet. Conversely, real news appears throughout the day. Therefore, we believe that these features are important signs in addition to text content to predict whether news is real or not.

Pretrained Models

Due to the rapid development of deep learning, a multitude of pretrained language models are now available for Vietnamese, categorized as multilingual and monolingual. Because this dataset is a collection of Vietnamese news and we found that monolingual models outperform multilingual models on this dataset⁸, we decided to use some monolingual models for this language, ViBERT, vELECTRA¹⁴ and PhoBERT¹¹, for this task. This list below is a summarized introduction of them:

PhoBERT^a: This model is pretrained by a masked language modeling task on 20 GB Vietnamese data, inherited from the original BERT¹⁵. In this pretraining task, PhoBERT is inherited from RoBERTA¹⁶, which cuts down the Next Sentence Prediction pretraining task from the original BERT for better performance.

^a<https://huggingface.co/vinai/phobert-base>

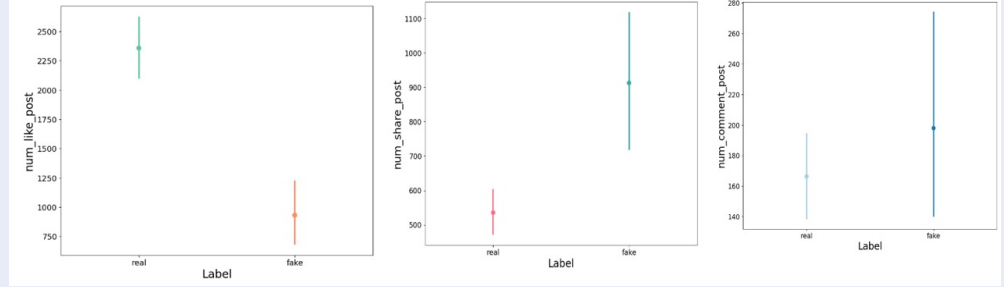


Figure 2: (a) Number of likes between real and fake news (b) Number of shares between real and fake news (c) Number of comments between real and fake news



Figure 1: (a) Fake news’ word cloud (b) Real news’ word cloud.

vELECTRA^b: This model is inherited from the original ELECTRA architecture⁴ and pretrained on the 60 GB Vietnamese language from NewsCorpus^c and OscarCorpus^d. Different from previous BERT-based architectures, this model uses a replaced token detection pretraining task. In this task, there are two neural networks trained simultaneously, a discriminator and a generator. The generator performs a masked language modeling (MLM) task, which chooses some random tokens to be replaced by the [MASK] token.

^b<https://huggingface.co/FPTAI/velectra-base-discriminator-case>

^c<https://github.com/binhvq/news-corpus>

^d<https://oscar-corpus.com/>

Then, the generator produces output tokens at the position of the masked-out tokens. The discriminator will try to determine whether those tokens have been replaced or come from the original sentence. The experimental results indicated that this pretraining task outperformed the MLM task of BERT.

ViBERT^e: This model is pretrained on a 10 GB Vietnamese corpus from preprocessed online newspapers.

Fine-tuning approach

In this section, we present our fine-tuning approaches with a pretrained language model. We recognized that many teams in 2020 VLSP competitions used transfer learning methods, but most of them fine-tuned on textual features only¹⁷ and ignored the other features, while determining fake news should be based on not only news content but also several components. Therefore, we decided to fine-tune them on text incorporating timestamp and reaction features by verbal words. For example, we have written sample sentences as follows: “This post was written on Friday, April 3rd at 8 PM. There are 19447 likes, 32 comments and 50 shares from this post” (in Vietnamese) to describe those features, and then we merge them with the original sentences to make new ones. We illustrated this process in Figure 3.

Our implementation is based on fine-tuning ideas proposed by¹⁵ and¹⁸. BERT-based models consist of 12 Transformer layers in the encoders, and the hidden size of each vector is 768. To fine-tune these architectures for classification, we must add the special tokens named [CLS] and [SEP] to the text input. The [CLS] token is encoded as contextual representation of the input, and the [SEP] is used as separation between sentences in the entire input string. Then, these representations from one or several layers are fed into a

^e<https://huggingface.co/FPTAI/vibert-base-cased>

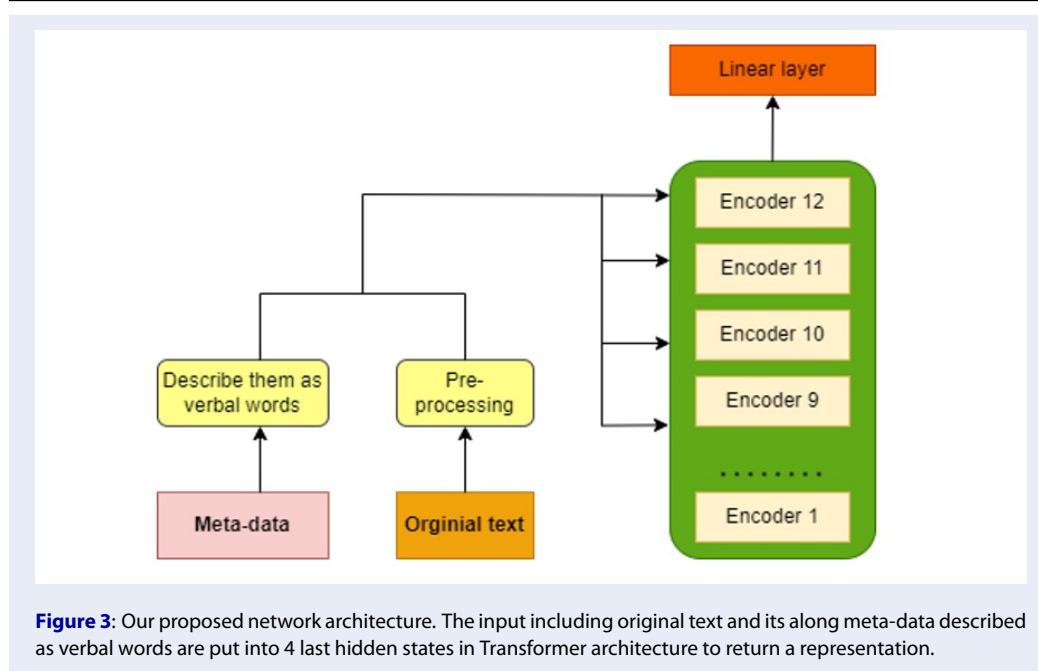


Figure 3: Our proposed network architecture. The input including original text and its along meta-data described as verbal words are put into 4 last hidden states in Transformer architecture to return a representation.

fully connected layer which contains a Softmax function to output probability of each label. In the classification part, we applied many advanced fine-tuning techniques: warm-up learning rate and unfreeze layers gradually in each training epoch.

According to previous works, we observed that lower layers in BERT architectures contain surface features and general information, and the information gradually obtains more granularity and complex semantic features in higher layers^{19,20}. Therefore, concatenating multiple layers together ensures that our information is both general and granular. In this case, we decided to concatenate the four last layers because it helps model obtain the best classification performance¹⁸.

Another factor that we need to consider is the text input's length. Because BERT-based architecture's input has a maximum length of 512 tokens. Moreover, we can only set the maximum length of 256 tokens due to computational computation. Therefore, we applied the "head tail" truncation method, which keeps the first 64 tokens and last 192 tokens because the key information usually comes from the beginning and ending of the entire text, which provides many meaningful insights.

We also intended to combine the image feature with the other features, including text, number of likes, comments and shares, into the model. However, the huge shortage of this information (Table 2) lets us give up them.

Training and testing procedures

In the training procedure, we fine-tune the model by feeding the encoded [CLS] token of the input information and use the concatenated representations from the four last hidden layers to a fully connected layer on top of BERT. During training, we use the binary cross entropy with logits loss (BCEWithLogits) to calculate the difference between the predicted model output's logits and the actual label. In the gradual unfreezing method, we first freeze all layers and then fine-tune them after a few epochs. This method ensures that gradients are not updated at the first epoch because the low-level layers already have generally learned features; therefore, we will avoid overfitting.

In the testing phase, the saved model outputs distributional probabilities of labels. Because the test data is unlabeled, we must submit the output probabilities to the competition's system due to their submission regulation.

EXPERIMENTS

Experimental configuration

We conducted experiments on Google Colab Pro with Tesla K80 GPU configuration, 13 GB RAM, and Intel Xeon CPU @ 2.2 GHz. We use the PyTorch^f library in Python programming language. We choose a batch size of 32 to save computational cost and AdamW

^f<https://pytorch.org/>

as the optimizer. The learning rate is selected in the range [1e-5, 1e-4]. All of our results are conducted on the private test set.

Evaluation Metrics

We use the Area Under the Receiver Operating Characteristic Curve for evaluation (AU-ROC). This is a performance measurement for classification at various thresholds. ROC is a probability curve, and AUC represents the degree or measure of separability. It tells how much the model can distinguish classes.

EXPERIMENTAL RESULTS

Our experimental results are reported by AUC score metrics in Tables 3, 4, 5 and 6. We will describe them in more detail below.

Table 3 shows that we applied some traditional machine learning models on meta-data only to investigate the effectiveness on these features. We achieved approximately 0.7 AUC score, which is a good performance on numerical features for fake news detection, with Linear SVM model.

Table 3: The performance of traditional machine learning algorithms on meta-data only

Method	AUC
Logistic Regression	0.6948
Linear SVM	0.7029
LightGBM	0.6292
Catboost	0.5602
Random Forest	0.6014

As shown in Table 4, we trained the model under two circumstances: trained on text only and text incorporated with meta-data. The results show that the performance of all pretrained models increases significantly when merging verbal words that describe all meta-data to the input sequence. In particular, vELECTRA achieved the best performance.

To make it clearer, we also conducted experiments on incorporating text with each feature to investigate which feature contributes best to the model's performance. Furthermore, the values of meta-data will also be different for each filling value (mean and median), so we want to observe how strong the classification performance is with each different filling method. Many of these experimental results prevailed the top teams' ones. As shown in Table 5, we found that the model always increased its accuracy regardless of features merged with original text as well as filling methods.

Therefore, we have achieved state-of-the-art results on this dataset. We have summarized all previous results in Table 6, borrowed from the contest's organizers¹⁷ (★).

DISCUSSION

Since there are many missing values appearing in this dataset and if we did not preprocess them carefully, they would reduce our model's predictive performance. In preprocessing steps, we removed the clear outliers, and then the distribution of data would be less skewed. Therefore, the mean value is a suitable one to impute those missing data. Poulos & Valle (2018) also addressed that the mean value is important and preferred for quantitative features for supervised learning²¹.

Following the experimental results in Table 3, although those methods applied to meta-data are not as effective as textual features, they could contain some predictive information instead of randomly guessing for this task because their performances are larger than 0.5 AUC score. Therefore, when we concatenated these features with text, it would help the new text contain more meaningful information. The results shown in Tables 4 and 5 prove that this is true, and the model always outperforms when we incorporate information about the distribution of reactions with news content. Additionally, we can capture the patterns that (1) fake news usually has more shares or even comments and (2) there are likely more likes coming from real news rather than fake news, which is based on not only this dataset but also many other datasets, such as the PolitiFact dataset and Gossip-Cop dataset¹⁴. That is the interesting point to study users' behaviors on social media, which can help exploit more information to detect disinformation more effectively.

CONCLUSION

To recapitulate, we have conducted effective methods and fine-tuning strategies to detect untrustworthy news: fine-tuning the ViBERT, PhoBERT and vELECTRA models combined with numeric features by verbal words. Our best result belongs to vELECTRA when it achieved 0.9575 AUC score on this dataset. Through many experiments, we observed that in addition to the text data, we can leverage other features for classification because they may contain key information to help the model predict more accurately.

Because identifying the reliability of the news is even difficult for humans, it can be determined not only on the news' content but also on related knowledge about

Table 4: The performance of models on original text and text combined with meta-data.

Pretrained model	learning rate	epochs	AUC
ViBERT	3e-05	6	0.9324
vELECTRA	1e-05	4	0.9390
PhoBERT	6e-05	6	0.9259
ViBERT + meta-data	3e-05	6	0.9380
vELECTRA + meta-data	1e-05	4	0.9575
PhoBERT + meta-data	6e-05	6	0.9519

Table 5: Experimental results of the v-ELECTRA – best model when incorporating different features and filling methods.

Mean filling		Median filling	
Feature	AUC	Feature	AUC
Num like post	0.9512	Num like post	0.9448
Num share post	0.9496	Num share post	0.9439
Num comment post	0.9492	Num comment post	0.9429
Timestamp	0.9422	Timestamp	0.9421
All	0.9575	All	0.9450

Table 6: Results comparison between our best model with other teams

Team/Author	AUC	Model
Ours	0.9575	vELECTRA + meta-data
(Pham et al., 2021)	0.9538	PhoBERT + CNN + TF-IDF
Kurtosis (★)	0.9521	TF-IDF + SVD; PhoBERT embedding
NLP BK (★)	0.9513	Bert4news + PhoBERT + XLMR.
SunBear (★)	0.9462	RoBERTa + MLP
uit kt (★)	0.9452	PhoBERT + BERT4news
Toyo-aime (★)	0.9449	CNN + BERT
Parabol (★)	0.9397	x
ZaloTeam (★)	0.9378	vELECTRA
UIT VLTN (★)	0.8102	x

users’ reactions. Thus, in the future, we propose to create a more elaborate dataset containing sources of news. The sources and authors that have a high percentage of fake news will be penalized strongly by the model. In contrast, in news that comes from accredited sources, authors will be more likely to be considered trustworthy news. Moreover, we want that the dataset will contain more images because we can take advantage of them since “a picture is worth a thousand words”, and images are the key pieces of evidence for verifying news. In addition, fake news problems

exist not only in Vietnamese but also in other languages, and multilingual methods are worth considering for a future of less fake news. Therefore, in the future, we are going to improve the performance by multimodal and multilingual approaches.

ACKNOWLEDGEMENT

None

AUTHOR CONTRIBUTION

Khoa Pham Dang and Dang Van Thin discussed with each other and proposed the main ideas of this research. Khoa Pham Dang implemented, conducted the experiments and wrote the paper. Ngan Luu-Thuy Nguyen discusses the ideas and contributed to writing the paper. All authors approved the final version.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

ABBREVIATIONS

None

REFERENCES

- Allcott H, Gentzkow M. Social media and fake news in the 2016 election. *J Econ Perspect*. 2017;31(2):211-36; Available from: <https://doi.org/10.1257/jep.31.2.211>.
- Orso D, Federici N, Copetti R, Vetrugno L, Bove T. Infodemic and the spread of fake news in the COVID-19-era. *Eur J Emerg Med*. 2020;May 5(5):327-8; PMID: 32332201. Available from: <https://doi.org/10.1097/MEJ.0000000000000713>.
- Agarwal A, Meel P. Stacked bi-LSTM with attention and contextual BERT embeddings for fake news analysis 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India. Vol. 2021; 2021. p. 233-7; Available from: <https://doi.org/10.1109/ICACCS51430.2021.9441715>.
- Clark K, Luong M-T, Le QV, Manning D. Electra: pretraining text encoders as discriminators rather than generators. In: Proceedings of the international conference on learning representations (ICLR). Vol. 2020; 2020;.
- Qazvinian V, Rosengren E, Radev D, Mei Q. Rumor has it: identifying misinformation in microblogs. In: Proceedings of the 2011 conference on empirical methods in natural language processing; 2011. p. 1589-99;.
- Thanh KNT, Van KN. Reintel challenge 2020: exploiting transfer learning models for reliable intelligence identification on Vietnamese social network sites. In: Proceedings of the 7th international workshop on Vietnamese language and speech processing; 2020. p. 45-8;.
- Trinh HV, Bui TT, Nguyen TM, Dao HQ, Pham QH, Tran NN et al. Reintel challenge 2020: A comparative study of hybrid deep neural network for reliable intelligence identification on Vietnamese sns. In: Proceedings of the 7th international workshop on Vietnamese language and speech processing; 2021;.
- Tran T-H, Phan L, Nguyen T-S, Nguyen T-H. Leveraging transfer learning for reliable intelligence identification on Vietnamese snss (reintel). In: Proceedings of the 7th international workshop on Vietnamese language and speech processing; 2020. p. 45-8;.
- Nguyen MD, Pham QN. ReINTEL challenge 2020: A multi-modal ensemble model for detecting unreliable information on Vietnamese SNS. In: Proceedings of the 7th international workshop on Vietnamese language and speech processing 2020 dec; 2020. p. 40-4;.
- Hieu TN, Minh HCN, Van HT, Quoc V. Reintel challenge 2020: Vietnamese fake news detection using ensemble model with phobert embeddings. In: Proceedings of the 7th international workshop on Vietnamese language and speech processing; 2020. p. 1-5;.
- Nguyen DQ, Tuan Nguyen AT. PhoBERT: pre-trained language models for Vietnamese. *Find Assoc Comp Linguist EMNLP*. 2020;2020(v):1037-42). <https://doi.org/10.18653/v1/2020findings-emnlp92>;
- Vu T, Nguyen DQ, Nguyen DQ, Dras M, Johnson M. VnCoreNLP: A Vietnamese natural language processing toolkit. In: Proceedings of the annual conference of the north American Chapter of the Association for Computational Linguistics. Vol. 2018; 2018. p. 56-60; Available from: <https://doi.org/10.18653/v1/N18-5012>.
- Jin Z, Cao J, Zhang Y, Luo Jiebo. News verification by exploiting conflicting social viewpoints in microblogs. *AAAI*. 2016;30(1); Available from: <https://doi.org/10.1609/aaai.v30i1.10382>.
- Shu K, Mahudeswaran D, Wang S, Lee D, Liu H. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big Data*. 2020;8(3):171-88; PMID: 32491943. Available from: <https://doi.org/10.1089/big.2020.0062>.
- Devlin J, Chang M-W, Lee K, Toutanova K. BERT: pretraining of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 conference of the north American Chapter of the Association for Computational Linguistics: Human Language Technologies (Long and short papers). Vol. 1. MN: Association for Computational Linguistics; 2019. p. 4171-86, Minneapolis;.
- Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D et al. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*. 2019. 2019; Jul 26;.
- Le DT, Vu XS, To ND, Nguyen HQ, Nguyen TT, Le L et al. ReINTEL: A multimodal data challenge for responsible information identification on social network sites. In: Proceedings of the 7th international workshop on Vietnamese language and speech processing; 2020. p. 1-5;.
- Sun C, Qiu X, Xu Y, Huang X. How to fine-tune bert for text classification? In: Proceeding of China national conference on hinese computational linguistics; 2019. p. 194-206; Available from: https://doi.org/10.1007/978-3-030-32381-3_16.
- Jawahar G, Sagot B, Seddah D. What does BERT learn about the structure of language? In: Proceedings of the 57th annual meeting of the Association for Computational Linguistics. Italy: Association for Computational Linguistics; 2019. p. 3651-7, Florence; Available from: <https://doi.org/10.18653/v1/P19-1356>.
- Tenney I, Das D, Pavlick E. BERT rediscovers the classical NLP pipeline. In: Proceedings of the 57th annual meeting of the Association for Computational Linguistics; 2019. p. 4593-601; Available from: <https://doi.org/10.18653/v1/P19-1452>.
- Pham N-D, Le T-H, Do T-D, Vuong T-T, Vuong T-H, Ha Q-T. Vietnamese fake news detection based on hybrid transfer learning model and tf-idf. In: Proceeding of 13th international conference on knowledge and systems engineering (kse); 2021. p. 1-6; Available from: <https://doi.org/10.1109/KSE53942.2021.9648676>.