

Team baker at PAN: Enhancing Writing Style Change Detection with Virtual Softmax

Notebook for the PAN Lab at CLEF 2024

Bingpei Wu, Yong Han[†], Kai Yan and Haoliang Qi

Foshan University, Foshan, China

Abstract

This paper introduces the application of Virtual Softmax for the PAN 2024 multi author writing style analysis task. We found that tasks with the same topic are particularly challenging due to difficulties at classification boundaries. To address this problem, we integrated Virtual Softmax into the Transformer architecture to provide additional feature supervision, enhancing the recognition ability of model. Finally, we achieved F1 scores higher than the baseline method on the three tasks in the official test set.

Keywords

Style Change Detection, Pre-trained Model, Virtual Softmax

1. Introduction

Multi-Author Writing Style change detection involves identifying whether different paragraphs within the same document are written by different authors. [1] This technique holds significant importance in academic research and has extensive practical applications. For instance, in academia, style change detection can be used for detecting plagiarism in scholarly papers; in the publishing industry, it can help identify ghostwriting; and in the legal field, it can assist in verifying the authenticity of documents. Therefore, further research and enhancement of style change detection technology can not only elevate academic research but also provide robust technical support across various domains.

On this task PAN organized [2], Multi-Author Writing Style change detection is divided into three tasks:

- Task 1: The paragraphs of a document cover different topics.
- Task 2: The paragraphs of the document may cover different topics or the same topics.
- Task 3: All paragraphs in a document are on the same topic.

In this paper, we find that tasks with the same topic are difficult to identify for detecting author changes. The reason is that some data cause classification difficulties at classification boundaries. In this paper, we utilize the Virtual Softmax [3] to enlarge the inter-class margin and compress the intra-class distribution. Therefore, we integrate Virtual Softmax into a Roberta [4] architecture to enhance classification capabilities and provide feature supervision during training, thus enhancing the identification ability of features.

2. Related work

With the rise and improvement of large language model technology, the methods for handling complex tasks have undergone significant changes. Traditional work, such as the research by Gómez-Adorno et al. [5], relied on the design of stylometric features and the use of machine learning methods for

CLEF 2024: Conference and Labs of the Evaluation Forum, September 09–12, 2024, Grenoble, France

[†] Corresponding author

✉ wubingpei0819@gmail.com (B. Wu); hanyong2005@fosu.edu.cn (Y. Han); yankai@fosu.edu.cn (K. Yan);

qihaoliang@fosu.edu.cn (H. Qi)

ORCID 0009-0004-6281-4322 (B. Wu); 0000-0002-9416-2398 (Y. Han); 0000-0002-4960-7108 (K. Yan); 0000-0003-1321-5820 (H. Qi)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

prediction. However, the focus has now shifted towards fine-tuning large language models using various techniques. For instance, in PAN 2023, Ye et al. [6] work employed contrastive learning for supervised fine-tuning, achieving remarkable results.

In our work, we compared the performance of three pre-trained transformer-based models on Task 1 to select the most suitable base model. Additionally, we applied data augmentation techniques and used Virtual Softmax to enhance the model’s performance in handling samples near the classification boundaries.

3. Method

3.1. Network Architecture

The Transformer [7] architecture is technologically mature and includes well-developed models such as BERT [8], RoBERTa [4], and DistilBERT [9]. This architecture, pre-trained on a large corpus, has strong contextual understanding capabilities. We compared the F1 performance of models with the same parameters in task1, then chose RoBERTa as the base model, as shown in the Table1.

Table 1

Task1 Performance comparison

	Task 1
BERT	0.9155
DistilBERT	0.9164
RoBERTa	0.9741

In our work, we use the RoBERTa-based model as the encoder to process the input text. The input text paragraphs are first tokenized and then fed into the RoBERTa model for encoding. The pooled output of the [cls] token, represents the contextual features of the entire paragraph. This output is then passed through a Virtual Softmax layer. The model is trained using a cross-entropy loss function to perform our classification task. During training, the extracted paragraph features are fed into a Virtual Softmax layer, enabling the model to perform a three-class classification task. An additional class is introduced to provide feature supervision, which compresses the inter-class space of the other two classes, thereby enforcing stricter boundary constraints.

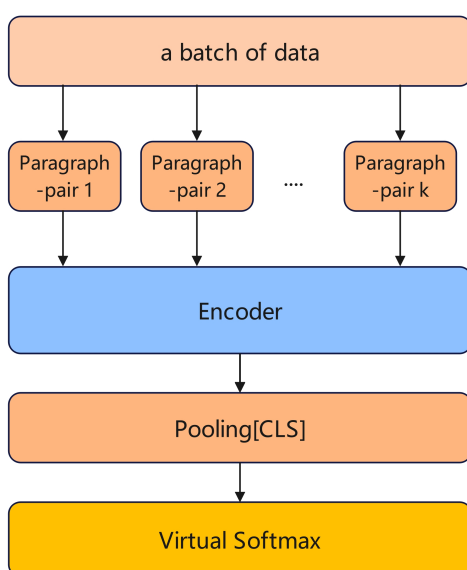


Figure 1: Model Architecture

3.2. Virtual Softmax

To enhance the model’s discriminative power, we integrate a Virtual Softmax layer. During the training phase, no additional processing is performed on the input data and Virtual Classes are directly added. During the evaluation phase, we choose the category with the highest probability among the non-virtual classes. These classes do not correspond to actual categories but are used to increase the complexity of the training process, thereby strengthening the model’s discriminative abilities. The core idea is to add noise by incorporating these virtual classes, forcing the model to generalize better when faced with real data. Specifically, for a given classification task, the additional injected classes introduce a new and tighter decision boundary for the original classes, compressing their inter-class distribution.

4. Experiments

4.1. Experiments setting

In this work, we select the RoBERTa-based model [4] for classification. The model consists of 12 layers and 12 attention heads, with a hidden size of 768. The maximum sequence length is set to 256, the learning rate is $1e-5$, and the batch size is 32. These experimental settings are consistent with the comparative experimental settings for BERT, RoBERTa, and DistilBERT.

4.2. Data preparation

In the data provided by PAN, three tasks, categorized by difficulty (task1-easy, task2-medium, task3-hard), were divided into a training set (70%), a validation set (15%), and a test set (15%). The training set for each difficulty consists of 4200 documents, consisting of multiple paragraphs. We connect two adjacent paragraphs in the same document using a separator token [cls] to form a sample, and label it whether the author of the sample text has changed.

Samples constructed according to the above method are made up of adjacent paragraphs. We extended the data set by linking discontinuous paragraphs together based on some logical judgment based on the number of authors and whether the document paragraphs changed. For example, if there is no author change for three consecutive paragraphs, the first paragraph and the third paragraph can form a new sample.

Table 2

Data processing and augmentation

	Task 1	Task 2	Task 3
train-set	11061	21906	19009
train-set-augment	13,504	26,695	24,548

4.3. Results

We submit the model to TIRA [10] for execution to get the final metrics for the model. Table 3 shows the F1 scores obtained by our model in the official test and validation set. In addition, the paper compares the performance of some methods in 2023, Chen et al. [11], Jacobo et al. [12].

In task1 and task2, authors can be distinguished by capturing the characteristics of the topic. Due to the architectural limitations of Transformer, style changes and context dependencies cannot be fully captured. This results in a low F1 score in task3, where each pair of paragraphs has the same topic and keywords.

Table 3
validation set result

Approach	Task 1	Task 2	Task 3
Our method	0.976	0.816	0.770
Chen et al.	0.914	0.820	0.676
Jacobo et al.	0.793	0.591	0.498
Baseline Predict 1	0.466	0.343	0.320
Baseline Predict 0	0.112	0.323	0.346

5. Conclusion

In this paper, we presented a RoBERTa-based model enhanced with Virtual Softmax for detecting style changes in multi-author documents. Our approach showed significant promise, particularly in the more challenging scenarios where documents share the same topic throughout. By injecting additional classes, we were able to improve the model’s ability to distinguish between different authors, thereby enhancing the robustness and accuracy of style change detection.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No.62276064).

References

- [1] E. Zangerle, M. Mayerl, M. Potthast, B. Stein, Overview of the Multi-Author Writing Style Analysis Task at PAN 2024, in: G. Faggioli, N. Ferro, P. Galuščáková, A. G. S. de Herrera (Eds.), Working Notes of CLEF 2024 - Conference and Labs of the Evaluation Forum, CEUR-WS.org, 2024.
- [2] J. Bevendorff, X. B. Casals, B. Chulvi, D. Dementieva, A. Elnagar, D. Freitag, M. Fröbe, D. Korenčić, M. Mayerl, A. Mukherjee, A. Panchenko, M. Potthast, F. Rangel, P. Rosso, A. Smirnova, E. Stamatatos, B. Stein, M. Taulé, D. Ustalov, M. Wiegmann, E. Zangerle, Overview of PAN 2024: Multi-Author Writing Style Analysis, Multilingual Text Detoxification, Oppositional Thinking Analysis, and Generative AI Authorship Verification, in: L. Goeuriot, P. Mulhem, G. Quénot, D. Schwab, L. Soulier, G. M. D. Nunzio, P. Galuščáková, A. G. S. de Herrera, G. Faggioli, N. Ferro (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Fifteenth International Conference of the CLEF Association (CLEF 2024), Lecture Notes in Computer Science, Springer, Berlin Heidelberg New York, 2024.
- [3] B. Chen, W. Deng, H. Shen, Virtual class enhanced discriminative embedding learning, *Advances in Neural Information Processing Systems* 31 (2018).
- [4] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach, *arXiv preprint arXiv:1907.11692* (2019).
- [5] H. Gómez-Adorno, J.-P. Posadas-Duran, G. Ríos-Toledo, G. Sidorov, G. Sierra, Stylometry-based approach for detecting writing style changes in literary texts, *Computación y Sistemas* 22 (2018) 47–53.
- [6] Z. Ye, C. Zhong, H. Qi, Y. Han, Supervised contrastive learning for multi-author writing style analysis, in: Conference and Labs of the Evaluation Forum (CLEF), 2023.
- [7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, *Advances in neural information processing systems* 30 (2017).
- [8] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, *arXiv preprint arXiv:1810.04805* (2018).
- [9] V. Sanh, L. Debut, J. Chaumond, T. Wolf, Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter, *arXiv preprint arXiv:1910.01108* (2019).

- [10] M. Fröbe, M. Wiegmann, N. Kolyada, B. Grahm, T. Elstner, F. Loebe, M. Hagen, B. Stein, M. Potthast, Continuous Integration for Reproducible Shared Tasks with TIRA.io, in: J. Kamps, L. Goeuriot, F. Crestani, M. Maistro, H. Joho, B. Davis, C. Gurrin, U. Kruschwitz, A. Caputo (Eds.), *Advances in Information Retrieval. 45th European Conference on IR Research (ECIR 2023)*, Lecture Notes in Computer Science, Springer, Berlin Heidelberg New York, 2023, pp. 236–241. doi:10.1007/978-3-031-28241-6_20.
- [11] H. Chen, Z. Han, Z. Li, Y. Han, A writing style embedding based on contrastive learning for multi-author writing style analysis, in: *Conference and Labs of the Evaluation Forum (CLEF)*, 2023.
- [12] G. X. Jacobo, V. Dehesa-Corona, A. D. Rojas-Reyes, H. Gómez-Adorno, Authorship verification machine learning methods for style change detection in texts, in: *Conference and Labs of the Evaluation Forum (CLEF)*, 2023.