Team NYCU-NLP at PAN 2024: Integrating Transformers with Similarity Adjustments for Multi-Author Writing Style Analysis

Notebook for the PAN Lab at CLEF 2024

Tzu-Mi Lin, Yu-Hsin Wu and Lung-Hao Lee*

Institute of Artificial Intelligence Innovation, National Yang Ming Chiao Tung University, Taiwan

Abstract

This paper describes our NYCU-NLP system design for multi-author writing style analysis tasks of the PAN Lab at CLEF 2024. We propose a unified architecture integrating transformer-based models with similarity adjustments to identify author switches within a given multi-author document. We first fine-tune the RoBERTa, DeBERTa and ERNIE transformers to detect differences in writing style in two given paragraphs. The output prediction is then determined by the ensemble mechanism. We also use similarity adjustments to further enhance multi-author analysis performance. The experimental data contains three difficulty levels to reflect simultaneous changes of authorship and topic. Our submission achieved a macro F1-score of 0.964, 0.857 and 0.863 respectively for the easy, medium and hard levels, ranking first and second, respectively for hard and medium levels out of 16 and 17 participating teams.

Keywords

Pre-trained Language Models, Embedding Similarity, Authorship Analysis, Plagiarism Detection

1. Introduction

The PAN Lab hosts a series of shared tasks for digital text forensics [1]. Following the achievements of the past Style Change Detection (SCD) tasks at the PAN Lab [2, 3], the goal of this multi-author writing analysis task seeks to identify all positions of writing style change at the paragraph level within a multi-authored document. Given a single document combined from separate comments by different users from the Reddit, the developed system should determine at which positions the author changes at three levels of difficulty: 1) Easy: the document contains multiple paragraphs on multiple topics; 2) Medium: the paragraphs in the document contains fewer topics; and 3) Hard: the document consists of multiple paragraphs on a single topic. All documents may contain an arbitrary number of style changes, which only occur between paragraphs.

This paper describes our developed **NYCU-NLP** (**N**ational **Y**ang Ming **C**hiao Tung **U**niversity, **N**atural **L**anguage **P**rocessing Lab) system. Our solution explores the use of three pre-trained transformers: RoBERTa, DeBERTa and ERNIE, and then fine-tunes the downstream classification task for the detection of changes to writing style. Finally, the system output is assembled using a majority voting-based assembly mechanism. We also take advantage of the property that sentences belonging to the same topic show greater similarity in the vector semantics space. We use the embedding similarity adjustments to enhance prediction performance at easy and medium levels which include paragraphs on different topics. Our final submission received macro F1-scores of 0.964, 0.857 and 0.863 respectively at the easy, medium and hard levels. These results ranked our method first and second, respectively for the hard and medium levels, out of 16 and 17 participating teams.

The rest of this paper is organized as follows. Section 2 reviews related studies. Section 3 describes our proposed NYCU-NLP system. Section 4 presents evaluation results and performance comparisons.

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

*Corresponding author.

ltmdegf4.ii12@nycu.edu.tw (T. Lin); amanda901014@gmail.com (Y. Wu); lhlee@nycu.edu.tw (L. Lee)

0000-0003-0472-7429 (L. Lee)

Conclusions are finally drawn in Section 5.

2. Related Work

The BERT transformer was used as the paragraph representation to train a random forest classifier for the SCD task [4]. Siamese neural networks were used to measure the paragraph similarities and identify authorship changes [5]. Individual transformers were trained independently and then assembled together for the final authorship change prediction [6]. The SCD task was regarded as a natural language inference task and solved using the DeBRETaV3 transformer [7]. A prompt-based approach was used to train a transformer model for the SCD task[8]. RoBERTa, BERT, and ELECTRA transformers were combined with a binary classification layer to solve the SCD task [9]. The SCD task was also regarded as an authorship verification problem based on the term-document matrix [10]. The mT0-x1 was used as the based teacher model to train the smaller student model based on the knowledge distillation mechanism [11]. A comparative learning method was presented to train the DeBERTa transformer to ensure paragraphs written by the same author are close in the semantic space [12].

In summary, using transformer-based models usually obtained promising results in the previous SCD tasks. Therefore, this motivates us to explore how to use transformers more effectively to solve the multi-author writing style analysis task at PAN-2024.

3. The NYCU-NLP System

Figure 1 shows our system architecture integrating transformers for multi-author writing style analysis, comprised of three main parts: 1) pre-trained transformers; 2) an assembly mechanism; and 3) similarity adjustments.

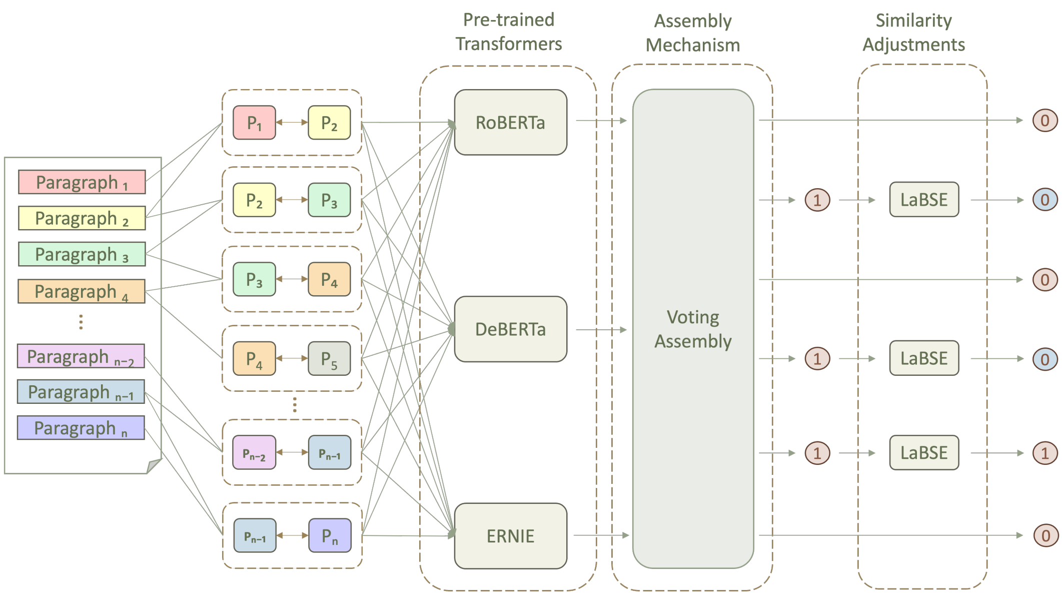

Figure 1: Our proposed NYCU-NLP system architecture

We first select the following transformers for multi-author writing style analysis:

- a Robust optimized BERT pre-training approach (RoBERTa) [13]
  RoBERTa enhances BERT [14] by removing the next sentence prediction objective that simplifies the training process, and using a dynamic masking strategy that improves model robustness. Furthermore, RoBERTa benefits from training with significantly larger batch sizes, enhancing the stability and effectiveness of the training process. These modifications result in a more robust

pre-trained language model that achieves superior performance on various natural language processing tasks.

- Decoding-enhanced BERT with disentangled attention (DeBERTa) [15]
DeBERTa improves BERT [14] by using a disentangled attention mechanism and an enhanced mask decoder. Each word is represented using content and position vectors and then disentangled matrices are used to compute attention weights. In the enhanced mask decoder architecture, absolute positions are used to predict the masked tokens for model pre-training.
- Enhanced Representation through Knowledge Integration (ERNIE) [16]
Inspired by the masking strategy of BERT [14], ERNIE is designed to learn language representations by entity-level masking and phrase-level masking. ERNIE 2.0 is an advanced version of ERNIE [17], which uses continuous multitask learning and a variety of pre-training tasks to enhance language comprehension. A continuous learning methodology is used to progressively integrate multiple tasks, which allows the model to proceed without forgetting what it has learned previously. In addition, ERNIE 2.0 proposes several new pre-training tasks, including word-aware, structure-aware, and semantic-aware tasks to respectively capture lexical information, syntactic information, and semantic information.

We fine-tune the language model of the individual pre-trained transformer and connected Multi-Layer Perceptron (MLP) as a classifier. Each pair of consecutive paragraphs is used for fine-tuning, along with its labeled classes (where '1' means change and otherwise '0'). We then use a voting-based assembly mechanism [18], which each transformer model makes an independent classification (i.e., a vote 0 or 1) for each testing instance. The final system output is determined by a majority of votes.

We suggest that two paragraphs with a similar topic should obtain a higher embedding similarity. Therefore, a multilingual LaBSE [19] embedding is used to represent each paragraph as a semantic vector. We then measure the cosine similarity between two given paragraph embedding vectors. If the similarity exceeds a predefined threshold, the topics of the two paragraphs should have a higher degree of similarity. We modify the assembly prediction from 1 (change) to 0 based on an assumption that paragraphs with similar topics usually reflect no change of author if the cosine similarity exceeds the threshold. In addition, since the paragraphs of a document at the easy and medium levels may contain a variety of topics, we only adopt this similarity adjustment mechanism at these two levels.

4. Evaluation

4.1. Data

The experimental datasets were mainly provided by task organizers [20]. Each level has 4,200 documents for model training and 900 documents for system validation. We also use additional 4,2000 documents each from the SCD-2023 task [3] to fine-tune the transformers for the medium and hard levels.

4.2. Settings

The pre-trained RoBERTa[1], DeBERTa[2], and ERNIE 2.0[3] models were downloaded from HuggingFace [21]. All models were fine-tuned on a server using a Nvidia Titan RTX GPU (24GB memory). The hyper-parameter values were optimized as follows: maximum sequence length of 256; learning rate 0.00005; dropout 0.25; epoch 10 and batch size 60. The LaBSE [4] was downloaded from TensorFlow Hub and the similarity adjustment threshold was set to 0.8. The system was deployed on the TIRA platform [22] to evaluate performance on the various difficulty levels using the macro-averaging F1-score.

[1]https://huggingface.co/roberta-base
[2]https://huggingface.co/microsoft/deberta-base
[3]https://huggingface.co/nghuyong/ernie-2.0-base-en
[4]https://tfhub.dev/google/LaBSE.

4.3. Results

Table 1 shows the validation set results. Among individual transformer models, DeBERTa-v1 outperformed the other models at the easy and hard levels. At the medium level, ERNIE 2.0 outperformed RoBERTa and DeBERTa. Our NYCU-NLP system used the assembly mechanism and similarity adjustments to obtain the best detection performance.

Table 2 shows the test set results. Our NYCU-NLP system significantly outperformed the baseline prediction for 1 or 0. We achieved a macro-averaging F1-score of 0.964 (ranking ninth of 17 systems) at the easy level; while F1-scores of 0.857 and 0.863 respectively at the medium and hard levels ranked first and second of 17 and 16 participating systems.

Table 1
Results of transformer models on the validation set.

Approach	Easy level	Medium level	Hard level
RoBERTa	0.9435	0.8436	0.8423
DeBERTa	0.9584	0.8408	0.8567
ERNIE 2.0	0.955	0.8496	0.849
NYCU-NLP	0.9716	0.8626	0.8658

Table 2
Submission results on the test set.

Approach	Easy level	Medium level	Hard level
NYCU-NLP	0.964	0.857	0.863
Baseline Predict 1	0.466	0.343	0.320
Baseline Predict 0	0.112	0.323	0.346

5. Conclusions

This study describes the design, implementation and evaluation of our NYCU-NLP system for the multi-author writing style analysis task at PAN 2024. We selected pre-trained transformer models as the starting points and fine-tuned the corresponding downstream classification tasks. Our unified architecture used a voting-based assembly mechanism to determine final system detection. We also adopted embedding similarity to adjust the system output at the easy and medium levels. Our submitted system ranked first of 17 participating systems at the hard level and second of 16 systems at the medium level.

Acknowledgments

This study is partially supported by the National Science and Technology Council, Taiwan, under the grant NSTC 111-2628-E-A49-029-MY3. This work was also financially supported by the Co-creation Platform of the Industry Academia Innovation School, NYCU.

References

[1] 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: Experimental IR Meets Multilinguality,

Multimodality, and Interaction. Proceedings of the Fourteenth International Conference of the CLEF Association (CLEF 2024), Lecture Notes in Computer Science, Springer, Berlin Heidelberg New York, 2024.

[2] E. Zangerle, M. Mayerl, M. Potthast, B. Stein, Overview of the Style Change Detection Task at PAN 2022, in: G. Faggioli, N. Ferro, A. Hanbury, M. Potthast (Eds.), CLEF 2022 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2022. URL: https://ceur-ws.org/Vol-3180/paper-186.pdf.

[3] E. Zangerle, M. Mayerl, M. Potthast, B. Stein, Overview of the Multi-Author Writing Style Analysis Task at PAN 2023, in: M. Aliannejadi, G. Faggioli, N. Ferro, M. Vlachos (Eds.), CLEF 2023 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2023. URL: https://ceur-ws.org/Vol-3497/paper-201.pdf.

[4] A. Iyer, S. Vosoughi, Style Change Detection Using BERT—Notebook for PAN at CLEF 2020, in: L. Cappellato, C. Eickhoff, N. Ferro, A. Névéol (Eds.), CLEF 2020 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2020. URL: http://ceur-ws.org/Vol-2696/.

[5] S. Nath, Style change detection using Siamese neural networks—Notebook for PAN at CLEF 2021, in: G. Faggioli, N. Ferro, A. Joly, M. Maistro, F. Piroi (Eds.), CLEF 2021 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2021. URL: http://ceur-ws.org/Vol-2936/paper-183.pdf.

[6] T.-M. Lin, C.-Y. Chen, Y.-W. Tzeng, L.-H. Lee, Ensemble Pre-trained Transformer Models for Writing Style Change Detection, in: G. Faggioli, N. Ferro, A. Hanbury, M. Potthast (Eds.), CLEF 2022 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2022. URL: https://ceur-ws.org/Vol-3180/paper-210.pdf.

[7] I. E. Kucukkaya, U. Sahin, C. Toraman, ARC-NLP at PAN 23: Transition-Focused Natural Language Inference for Writing Style Detection, in: M. Aliannejadi, G. Faggioli, N. Ferro, M. Vlachos (Eds.), CLEF 2023 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2023. URL: https://ceur-ws.org/Vol-3497/paper-218.pdf.

[8] Z. Zhang, Z. Han, L. Kong, Style Change Detection based on Prompt, in: G. Faggioli, N. Ferro, A. Hanbury, M. Potthast (Eds.), CLEF 2022 Labs and Workshops, Notebook Papers, CEUR-WS.org, 2022. URL: https://ceur-ws.org/Vol-3180/paper-197.pdf.

[9] A. Hashemi, W. Shi, Enhancing Writing Style Change Detection using Transformer-based Models and Data Augmentation, in: M. Aliannejadi, G. Faggioli, N. Ferro, M. Vlachos (Eds.), Working Notes of CLEF 2023 - Conference and Labs of the Evaluation Forum, CEUR-WS.org, 2023, pp. 2613–2621. URL: https://ceur-ws.org/Vol-3497/paper-212.pdf.

[10] G. Jacobo, V. Dehesa, D. Rojas, H. Gómez-Adorno, Authorship verification machine learning methods for Style Change Detection in texts, in: M. Aliannejadi, G. Faggioli, N. Ferro, M. Vlachos (Eds.), Working Notes of CLEF 2023 - Conference and Labs of the Evaluation Forum, CEUR-WS.org, 2023, pp. 2652–2658. URL: https://ceur-ws.org/Vol-3497/paper-217.pdf.

[11] M. Huang, Z. Huang, L. Kong, Encoded Classifier Using Knowledge Distillation for Multi-Author Writing Style Analysis, in: M. Aliannejadi, G. Faggioli, N. Ferro, M. Vlachos (Eds.), Working Notes of CLEF 2023 - Conference and Labs of the Evaluation Forum, CEUR-WS.org, 2023, pp. 2629–2634. URL: https://ceur-ws.org/Vol-3497/paper-214.pdf.

[12] H. Chen, Z. Han, Z. Li, Y. Han, A Writing Style Embedding Based on Contrastive Learning for Multi-Author Writing Style Analysis, in: M. Aliannejadi, G. Faggioli, N. Ferro, M. Vlachos (Eds.), Working Notes of CLEF 2023 - Conference and Labs of the Evaluation Forum, CEUR-WS.org, 2023, pp. 2562–2567. URL: https://ceur-ws.org/Vol-3497/paper-206.pdf.

[13] 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 (2019). doi:https://doi.org/10.48550/arXiv.1907.11692.

[14] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, Proceedings of NAACL-HLT 2019 (2019) 4171–4186. doi:https://doi.org/10.48550/arXiv.1810.04805.

[15] P. He, X. Liu, J. Gao, W. Chen, Deberta: Decoding-enhanced bert with disentangled attention, International Conference on Learning Representations (2021). doi:https://doi.org/10.48550/arXiv.1907.11692.

[16] Y. Sun, S. Wang, Y. Li, S. Fen, H. Tian, H. Wu, H. Wang, Ernie 2.0: A continual pre-training framework for language understanding, arXiv preprint arXiv:1907.12412 (2019).
[17] Z. Zhang, X. Han, Z. Liu, M. S. Xin Jiang, Q. Liu, Enhanced language representation with informative entities, Proceedings of ACL 2019 (2019) 1441–1451. doi:https://doi.org/10.18653/v1/P19-1139.
[18] L.-H. Lee, Y.-S. Wang, C.-Y. Chen, L.-C. Yu, Ensemble multi-channel neural networks for scientific language editing evaluation, Institute of Electrical and Electronics Engineers Access (2021) 158540 – 158547. doi:10.1109/ACCESS.2021.3130042.
[19] F. Feng, Y. Yang, D. Cer, N. Arivazhagan, W. Wang, Language-agnostic bert sentence embedding, Proceedings of ACL 2022 (2022) 878–891. doi:https://doi.org/10.18653/v1/2022.acl-long.62.
[20] 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.
[21] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. L. Scao, S. Gugger, M. Drame, Q. Lhoest, A. M. Rush, Huggingface's transformers: State-of-the-art natural language processing (2019). doi:https://doi.org/10.48550/arXiv.1910.03771.
[22] 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. URL: https://link.springer.com/chapter/10.1007/978-3-031-28241-6_20. doi:10.1007/978-3-031-28241-6_20.