Style Change Detection based on Prompt

Zhijie Zhang, Zhongyuan Han*, Leilei Kong

Foshan University, Foshan, China

Abstract

The Style Change Detection 2022 aims to identify text positions within a given multi-author document at which the author switches. The previous method for detecting the style change used the Pretrained Language Model BERT to extract the interaction features of paragraphs as author style change. However, the models like BERT have a problem of inconsistent train and test objectives as there is no Mask Language Model (MLM) task when fine-tuning the downstream. Moreover, the effectiveness of features certainly lacks interpretability. For the above two points, this paper proposes a method of Style Change Detection based on Prompt (SCDP), a novel way of interacting with building a manual Prompt to reflect the writing style between two adjacent paragraphs or sentences by utilizing a pre-trained MLM. Using the proposed SCDP, we can settle the problem of inconsistent train and test objectives and increase the interpretability of the model, revealing the possibility of using a Pretrained Language Model-based Prompt for Style Change Detection.

Keywords

Style Change Detection, Prompt, Mask Language Model

1. Introduction

In recent years, we have witnessed the importance of writing style change detection, which is related to problems from the fields of stylometry, intrinsic plagiarism detection, and authorship attribution[1][4]. The style change detection task aims to identify text positions within a given multi-author document at which the author switches. At PAN 2022 Style Change Detection[1], there is one more sentence-level task than PAN 2021. The difficulty level increased as the sentence level meant less information is available, making it more difficult to distinguish the author's writing style.

The previous method for detecting the style change like Zhang et al. [2], using the Pretrained Language Model BERT[6]; however, research[7] recently shows the problems of BERT that inconsistent train and test objectives. Moreover, Are the text interaction features extracted after fine-tuning the BERT features about writing style? We believe that its lacks certain interpretability.

In response to the above two points, we propose a method of Style Change Detection based on Prompt (SCDP), a novel way of interaction with building a manual Prompt (i.e. manual construction of templates) to reflect the writing style between two adjacent paragraphs or sentences by utilizing a pre-trained MLM. Using the proposed SCDP, we can settle the problem of inconsistent train and test objectives and increase the interpretability of the model, revealing the possibility of using a Pretrained Language Model-based Prompt for Style Change Detection.

CLEF 2022 – Conference and Labs of the Evaluation Forum, September 5–8, 2022, Bologna, Italy

EMAIL: zhangzhijie5454@gmail.com (A. 1); hanzhongyuan@gmail.com (A. 2)(*corresponding author); kongleilei@fosu.edu.cn (A. 3)
ORCID: 0000-0002-4854-0618 (A. 1); 0000-0001-8960-9872 (A. 2); 0000-0002-4636-3507 (A. 3)

2. Style Change Detection

This section briefly describes the tasks and datasets.

- **Tasks:** Style Change Detection 2022[1] was divided into three tasks. Specifically, Style Change Basic (Task1): For a text written by two authors that only contains a single style change, find this change's position. Style Change Advanced (Task2): For a text written by two or more authors, find all positions of writing style change. Style Change Real-World (Task3): For a text written by two or more authors, find all positions of writing style change, where style changes now occur between paragraphs and at the sentence level.

- **Datasets:** Style Change Detection 2022 released new datasets[5], which are based on user posts from various sites of the StackExchange network, covering different topics. Three datasets, including ground truth information, are provided (dataset1 for task 1, dataset2 for task 2, and dataset3 for task 3).

3. Method

In this section, we briefly outline our approach. We have mainly divided into tasks solution and SCDP model construction.

3.1. Tasks Solution

After analyzing three task definitions, we found that Task1 and Task2 were the same as Task2 and Task3 last year, and Task3 is a sentence-level detection. The difficulty level has increased because the sentence level means that less information is available, making it more difficult to distinguish the author's writing style.

For the solution of the tasks, we followed the previous year's approach[2] and used only one model to complete three tasks. Specifically, for Task1, we used the model to determine whether the writing style between adjacent paragraphs is the same. A similar goes for Task3 but replacing paragraphs with sentences. While Task2, like Task3 last year, labels the author's serial number from the first paragraph of each text. We referred to the Zhang[2], which converts the standard serial number labels into binary labels first, and then judges whether the author between label-related paragraphs is the same (i.e., paragraphs that could affect the label change, please refer to the Zhang[2] for details).

3.2. SCDP Model

Based on the above solution, we also need a model that can discriminate the writing style between two paragraphs or sentences, and we mainly modify it for the BERT model. Inspired by Brown[3], which has reformulated the different NLP tasks as fill-in-the-blanks problems by different prompts, we propose a prompt-based method using the template to resolve training and testing objective inconsistencies of the Pretrained Language Model (PLM).

As shown in Figure 1, the Mask Language Model (MLM) is built up by BERT and MLM Head (a network mapping the hidden size to the vocabulary size). After constructing the input sample by a template like ***They are the [MASK] writing style:*** $\boldsymbol{P_i}$ ***and*** $\boldsymbol{P_{i+1}}$, where is a placeholder to put paragraph, represents the number of the paragraphs and represents the predicted token, we send it to MLM and start to predict the [MASK] token. Then, MLM will output the prediction vocabulary scores, where the max scores will be mapped to token [same], and its id (namely token id after tokenizer) will be the final label if the style between the previous paragraph and next paragraph has kept, or the final label will be the id of token [different]. The cross-entropy loss is applied for training to classify vocabulary. During the evaluation phases, the label [same] and [different] will be mapped back to corresponding labels [0] and [1].

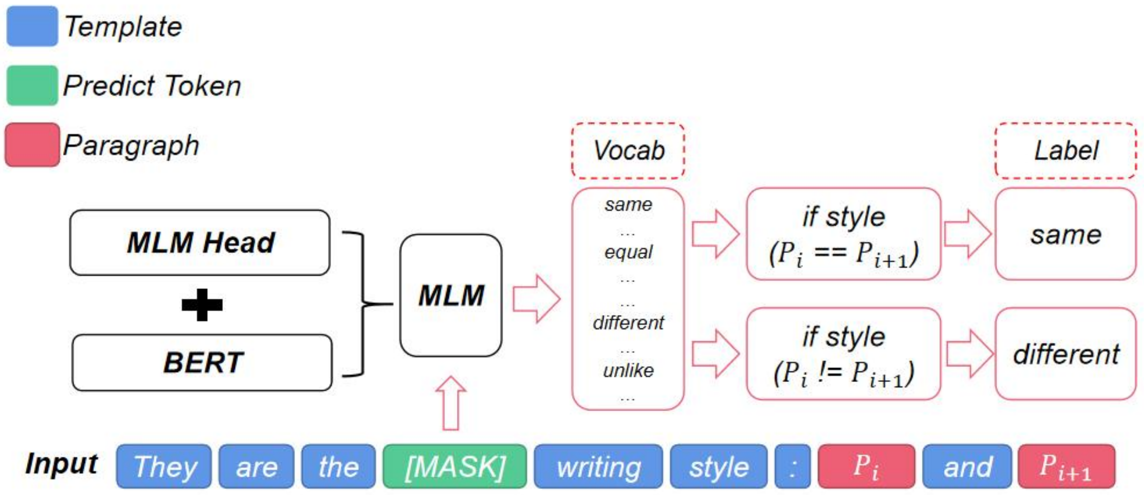

Figure 1: The illustration of SCDP.

Specifically, the blue squares is the template we have constructed, the red squares is the two input text paragraphs, and the green square predicts whether the writing style of the two texts has changed. In this structure, the token predicted by the model will have information about the writing style because MLM is a bi-directional language model that learns the context of the whole sentence through self-attention.

In this way, the acute problem of whether the extracted interaction features are stylistic features is avoided, and the introduction of the template also allows the model to understand better the problem we are trying to solve. Finally, it adds some interpretability to the model, at least to determine that the output of the model is based on the ***writing style*** of template.

4. Experimental setting

We conduct experiments on the RTX 3090 and use Pytorch implementation with CUDA to speed up training. The model BERT (bert-base-uncased) is utilized to build up MLM, which is equal to the pre-training phases of BERT. The experimental setting is shown in Table 1. Finally, evaluate the model on TIRA[8]. [1]

Table 1
Experimental setting

Batch size	Max Length	Learning Rate	Epoch	Report step
64	256	3e-5	3	125

5. Results

The trained model is evaluated on the validation and test sets, with the results shown in Table 2. To better measure Task2, two additional evaluations, Diarization Error Rate (DER) and Jaccard Error Rate (JER), were added to Task 2 this year as the labels of each text in Task 2 are interrelated.

The results show that the validation and test sets are similar, proving that the method is stable. Although it may not be as good as the traditional fine-tuning method, our approach overcomes the

[1] Our source code is publicly available at https://github.com/chigee54/SCDP

model's shortcomings with inconsistent training and test objectives. We will continue to explore how to use the prompt-based model to identify writing style better, believing that it can increase the interpretability of the model.

Table 2
The result of Datasets

Datasets	Task1.F1	Task2.F1	Task2.DER	Task2.JER	Task3.F1
Validation	0.70456	0.43098	0.71707	0.65023	0.66688
Test	0.71623	0.41741	0.71140	0.64441	0.65814

6. Conclusion

This paper proposes a method of Style Change Detection based on Prompt (SCDP), a novel way of interacting with building a manual Prompt to reflect the writing style between two adjacent paragraphs or sentences by utilizing a pre-trained MLM. Using the proposed SCDP, we can settle the problem of inconsistent train and test objectives and increase the interpretability of the model, revealing the possibility of using a Pretrained Language Model-based Prompt for Style Change Detection.

7. Acknowledgments

This work is supported by the Natural Science Foundation of Guangdong Province, China (No. 2022A1515011544).

8. References

[1] Eva Zangerle, Maximilian Mayerl, and Martin Potthast, and Benno Stein. Overview of the Style Change Detection Task at PAN 2022. In CLEF 2022 Labs and Workshops, Notebook Papers, CEUR Workshop Processings.

[2] Zhijie Zhang, Zhongyuan Han, Leilei Kong, et al. Style Change Detection Based On Writing Style Similarity—Notebook for PAN at CLEF 2021. In Guglielmo Faggioli et al., editors, CLEF 2021 Labs and Workshops, Notebook Papers, September 2021. CEUR-WS.org.

[3] Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. arXiv preprint arXiv:2005.14165.

[4] J. Bevendorff, B. Chulvi, E. Fersini, A. Heini, M. Kestemont, K. Kredens, M. Mayerl, R. Ortega, P. Pezik, M. Potthast, F. Rangel, P. Rosso, E. Stamatatos, B. Stein, M. Wiegmann, M. Wolska, and E. Zangerle. "Overview of PAN 2022: Authorship Verification,Profiling Irony and Stereotype Spreaders, and Style Change Detection," in 13th International Conference of the CLEF Association (CLEF 2022). Springer, 2022.

[5] E. Zangerle, M. Mayerl, M. Tschuggnall, M. Potthast, and B. Stein, " Pan22 authorship analysis: Style change detection," 2022. https://zenodo.org/record/6334245#.YiYCcXXMJ8x

[6] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

[7] Yang, Zhilin, et al. "Xlnet: Generalized autoregressive pretraining for language understanding." Advances in neural information processing systems 32 (2019).

[8] M. Potthast, T. Gollub, M. Wiegmann, and B. Stein, "TIRA Integrated Research Architecture," in Information Retrieval Evaluation in a Changing World, ser. The Information Retrieval Series, N. Ferro and C. Peters, Eds. Berlin Heidelberg New York: Springer, Sep. 2019.