SCL-DeBERTa: Multi-Author Writing Style Change Detection Enhanced by Supervised Contrastive Learning

Notebook for PAN at CLEF 2025

Kaichuan Lin, Chang Liu, Fuchuan Ye and Zhongyuan Han*

Foshan University, Foshan, China

Abstract

Multi-author writing style change detection, a core challenge in PAN@CLEF evaluations, requires precise localization of author transition points in collaboratively authored documents. This paper presents **SCL-DeBERTa**, novel framework that integrates **supervised contrastive learning** (SCL) with DeBERTa for fine-grained style boundary detection. in the PAN 2025 evaluation, Our team, **xxsu-team**, proposed a novel method that achieved F1 scores of **0.955 (ranking 3rd)**, **0.825 (ranking 1st)**, and **0.829 (ranking 2nd)** on the three subtasks, respectively, establishing a new benchmark for multi-author writing style analysis. This performance demonstrates significant advancements in feature extraction robustness and cross-task adaptability.

Keywords

PAN 2025, SCL-DeBERTa, DeBERTa, Supervised Contrastive Learning, Feature Disentanglement, Writing Style Change Detection,

1. Introduction

Writing style analysis, as a core technology in digital text forensics, aims to address the problem of author transition detection in multi-author mixed documents. Its primary goal is to achieve precise identification and localization of author switching boundaries by quantifying the stylistic differences between document sentence, making it a key research direction in the fields of digital text tracing and identity authentication. [1] This task has been continuously promoted by the PAN Lab at the CLEF conference. In this competition, the data is divided into three levels based on textual topic consistency: simple level (multi-topic documents, where topic differences can assist detection), intermediate level (limited topic variation requiring a focus on stylistic analysis), and difficult level (entirely reliant on pure stylistic feature recognition). Participants are required to use technologies such as deep learning to build models to solve related tasks.

To address this problem, we propose an innovative framework based on Supervised Contrastive Learning (SCL)[2] and DeBERTa[3]—**SCL-DeBERTa**. This model enhances the discriminative ability of stylistic features by introducing a supervised contrastive learning mechanism and effectively reduces topic interference through a feature decoupling module. In the PAN 2025 evaluation, our method achieved F1 scores of 0.953, 0.823, and 0.83 on the three subtasks, providing a new perspective and approach for multi-author writing style analysis tasks.

2. Related Work

The identification of writing styles in multi-author documents has emerged as a core task in digital text forensics, with numerous solutions based on pre-trained language models proposed in recent years. Early approaches primarily relied on traditional feature engineering, such as lexical statistical features

CLEF 2025 Working Notes, 9 - 12 September 2025, Madrid, Spain

*Corresponding author.

1441584221bruan@gmail.com (K. Lin); lc965024004@gmail.com (C. Liu); yfc2097639788@gmail.com (F. Ye); hanzhongyuan@gmail.com (Z. Han*)

0009-0001-7843-2888 (K. Lin); 0009-0000-0887-9273 (C. Liu); 0009-0002-1467-0358 (F. Ye); 0000-0001-8960-9872 (Z. Han*)

and syntactic pattern analysis, but these methods exhibited limited generalization ability in cross-topic scenarios.

With the rise of Transformer models, researchers began exploring the adaptability of different pre-trained architectures: Team Gladiators[4] systematically evaluated the performance of models such as ELECTRA, SqueezeBERT, and RoBERTa, finding that RoBERTa performed best in cross-topic tasks due to its dynamic masking strategy. However, its paragraph order reversal data augmentation strategy, while alleviating class imbalance issues, failed to effectively capture deep stylistic features. Team NYCU-NLP[5] innovatively integrated RoBERTa, DeBERTa, and ERNIE models, introducing a semantic similarity correction mechanism based on LaBSE embeddings to explore the effects of model ensemble and semantic enhancement on performance. This method achieved an F1 score of 0.863 on the hard task, but the multi-model ensemble incurred over three times the computational cost, and the manually set topic similarity threshold limited the method's generalization ability. Chen et al.[6] optimized DeBERTa's embedding space using supervised contrastive learning, forcing paragraph vectors from the same author to cluster together. Although this method improved sensitivity to stylistic features, it relied on a complex paragraph pair reconstruction strategy, resulting in over 30% noisy false negatives when the number of authors was unknown. Large model adaptation techniques: The xxsu-team[7] was the first to apply LLAMA-3-8B to this task, compressing model parameters through Low-Rank Adaptation (LoRA)[8]. Although it achieved an F1 score of 0.887 on the medium task, int8 quantization led to representation information loss, causing performance fluctuations.

In contrast, we adopted the lightweight DeBERTa as the encoder to embed data samples. Its disentangled attention mechanism[3] separates content and positional information modeling, demonstrating stronger topic independence in multi-author style analysis. Additionally, we utilized a supervised contrastive learning framework to optimize model parameters. In the official PAN 2025 evaluation, our system achieved F1 scores of 0.953, 0.823, and 0.83 on the easy, medium, and hard subtasks, respectively, providing a reliable solution for real-world scenarios such as lightweight academic co-authorship detection and legal document tampering identification.

3. Data Processing

The PAN25 writing style analysis evaluation task focuses on paragraph-level stylistic change detection in multi-author documents, requiring participating systems to precisely identify all writing style transition boundaries within documents under strictly constrained conditions of author identity and topic variation.

3.1. Data Sources and Structure

The dataset used in this study for stylistic change detection consists of multiple document pairs, with each document corresponding to two files:

- `problem-<id>.txt`: Contains the original text content.
- `truth-problem-<id>.json`: Contains the annotations for stylistic changes.

The annotation information for each document is stored as a binary array, indicating whether a stylistic change occurs between adjacent sentences:

$$\text{changes} = [c_1, c_2, \ldots, c_{n-1}] \quad \text{where} \quad c_i \in \{0, 1\} \tag{1}$$

Here, n represents the total number of sentences in the document. The overall distribution of the dataset is shown in Table 1.

Table 1
Dataset Statistics

Datasets	Easy		Medium		Hard	
	#documents	#para.	#documents	#para.	#documents	#para.
Training set	4200	52602	4200	63017	4200	55262
Validation set	900	11147	900	13659	900	11548

3.2. Sample Structure

Each generated training sample contains the following elements:

$$\text{sample} = \{\underbrace{\text{input\_ids}}_{\text{Token IDs}}, \underbrace{\text{attention\_mask}}_{\text{Attention Mask}}, \underbrace{\text{label}}_{\text{Stylistic Label}}, \underbrace{\text{doc\_id}}_{\text{Document Identifier}}, \underbrace{\text{pair\_idx}}_{\text{Sentence Pair Index}}\} \quad (2)$$

3.3. Support for Supervised Contrastive Learning

To support the supervised contrastive learning framework, additional document-level metadata is included in the samples:

- **doc_id**: A unique document identifier used to group sentence pairs from the same document.
- **pair_idx**: The position index of the sentence pair within the document.

This metadata enables the model to learn representations of stylistic consistency within a document, enhancing its ability to detect stylistic changes in long documents.

3.4. Error Handling Mechanism

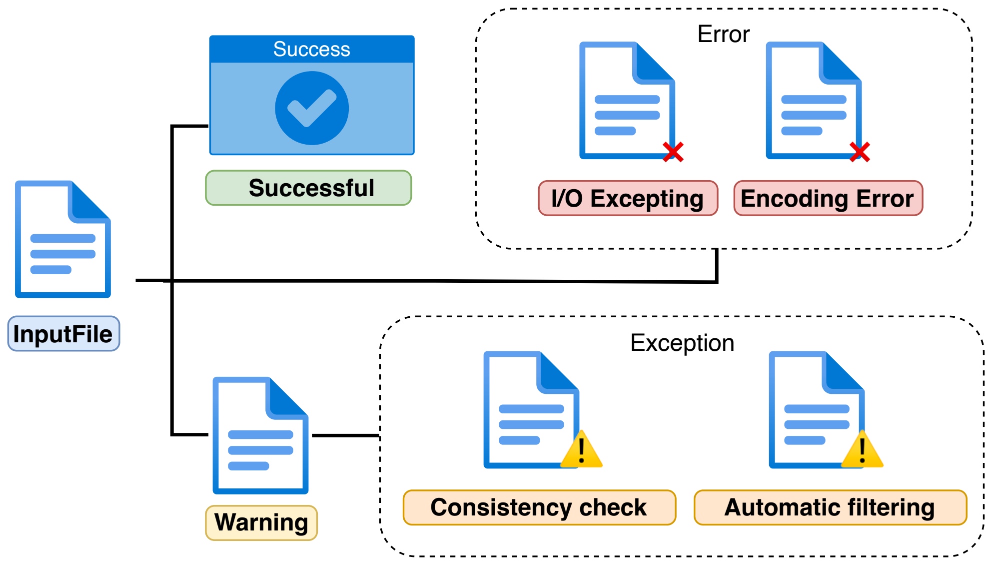

Figure 1: Exception/error procession

To ensure the robustness of data processing, the following error handling mechanisms are implemented:

- File I/O exception handling.
- Encoding error handling.

- Data consistency checks.
- Automatic filtering of invalid samples.

These mechanisms ensure that most usable data can still be processed effectively, even in cases of partial data corruption or non-standard formatting.

4. Method

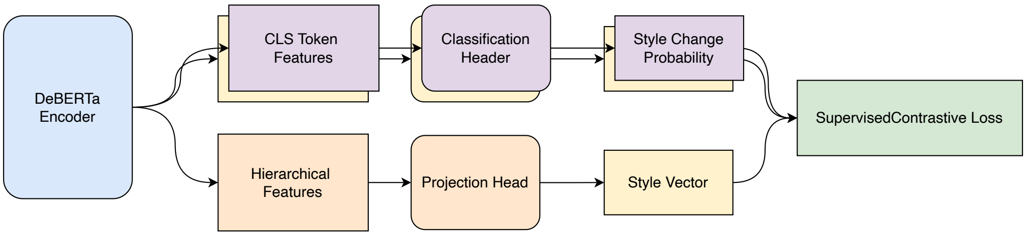

Figure 2: SCL-DeBERTa architecture

4.1. Task Definition

The multi-author writing style detection task, established by PAN Lab in CLEF conferences, requires precise localization of author transition points in collaboratively authored documents. Formally, given a document $D = \{p_1, p_2, \ldots, p_n\}$ composed of n paragraphs, the goal is to identify the boundary set $B = \{b_i \mid \text{author}(p_i) \neq \text{author}(p_{i+1})\}$.

4.2. SCL-DeBERTa Architecture

We propose **SCL-DeBERTa**, a novel framework integrating supervised contrastive learning with DeBERTa[3] encoder. As illustrated in Figure 2, the model employs dual-path processing:

$$\mathbf{h}_{\text{CLS}} = \text{DeBERTa}(\mathbf{x})[0] \tag{3}$$

$$\text{Classification head:} \quad \hat{y} = \sigma(\mathbf{W}_2 \cdot \text{ReLU}(\mathbf{W}_1 \mathbf{h}_{\text{CLS}})) \tag{4}$$

$$\text{Projection head:} \quad \mathbf{z} = \mathbf{W}_4 \cdot \text{ReLU}(\mathbf{W}_3 \mathbf{h}_{\text{CLS}}) \tag{5}$$

where $\mathbf{W}_1 \in \mathbb{R}^{256\times768}$, $\mathbf{W}_2 \in \mathbb{R}^{1\times256}$, $\mathbf{W}_3 \in \mathbb{R}^{256\times768}$, $\mathbf{W}_4 \in \mathbb{R}^{128\times256}$ are learnable parameters.

4.3. Supervised Contrastive Learning

We introduce a label-guided contrastive loss[9] to enhance style discrimination:

$$\mathcal{L}_{\text{scl}} = -\sum_{i=1}^{N} \frac{1}{|P(i)|} \sum_{p\in P(i)} \log \frac{\exp(\mathbf{z}_i \cdot \mathbf{z}_p/\tau)}{\sum_{k\neq i} \exp(\mathbf{z}_i \cdot \mathbf{z}_k/\tau)} \tag{6}$$

where $P(i) = \{p \mid y_p = y_i\}$ denotes positive samples sharing the same style label, and $\tau = 0.07$ is the temperature hyperparameter.

The joint optimization objective combines classification and contrastive losses:

$$\mathcal{L}_{\text{total}} = \underbrace{\text{BCE}(\hat{y}, y)}_{\mathcal{L}_{\text{cls}}} + \lambda \cdot \underbrace{\mathcal{L}_{\text{scl}}}_{\text{style regularization}} \tag{7}$$

with $\lambda = 0.3$ controlling the contrastive regularization strength.

5. Experiment

5.1. Dataset

The PAN 2025 multi-author writing style analysis evaluation task focuses on paragraph-level stylistic change detection, requiring participants to analysis the stylistic consistency of consecutive sentence pairs, participants must accurately identify all author transition boundaries. The task is designed to strictly control the synchronization of author identity and topic variation, ensuring that stylistic discrimination is not affected by topic shifts. Systems are evaluated across three difficulty levels based on the complexity of document topics.

1. **Easy Subtask**: The document topics are highly consistent, with frequent and obvious author transitions, and significant stylistic differences between paragraphs.
2. **Medium Subtask**: The document topics are moderately complex, with more subtle author transitions and smaller stylistic differences between paragraphs.
3. **Hard Subtask**: The document topics are highly mixed, with extremely subtle author transitions and almost imperceptible stylistic differences between paragraphs.

The PAN 2025 multi-author writing style analysis evaluation task centers on paragraph-level stylistic change detection, challenging participants to conduct pure stylistic change detection at the sentence level. By examining the stylistic consistency between consecutive sentence pairs, participants are required to precisely identify all author transition boundaries. The task is meticulously designed to enforce strict synchronization between author identity and topic variation, ensuring that stylistic discrimination remains unaffected by topic shifts. Systems are assessed across three difficulty levels, determined by the complexity of document topics.

5.2. Settings

In this study, we selected DeBERTa as the pre-trained model and utilized a supervised contrastive learning framework to optimize model parameters. Based on this foundation, we proposed *SCL-DeBERTa*, which enhances the ability of supervised contrastive learning to improve style differentiation. By combining classification and contrastive losses, the model achieves precise detection of author transition boundaries. The trained model was evaluated on three distinct task datasets, resulting in customized models tailored for each task. The hyperparameter settings used in this experiment are detailed in Table 2.

Table 2
Hyperparameter settings used in our experiments.

Hyperparameter	Value
max length	128
projection dim	128
dropout	0.1
batch size	32
max sequence length	512
initial learning rate	2×10^{-5}
epochs	10
temperature	0.07
weight decay	0.01

6. Results

6.1. Evaluation on PAN 2025 Benchmark

As shown in Table 3, *SCL-DeBERTa* achieved F1 scores of **0.955 (ranking 3rd)**, **0.825 (ranking 1st)**, and **0.829 (ranking 2nd)** on the three subtasks:

Table 3
Leaderboard results on PAN 2025 benchmark

Team	Approach	Task1 F1	Task2 F1	Task3 F1
wqd	felt-bronze	**0.958**	0.823	**0.830**
better-call-claude	medium-golfer	0.929	0.815	0.731
better-call-claude	senile-fortress	0.922	0.798	0.682
cornell-1	Ensembled-BertStyleNN	0.909	0.793	0.698
xxsu-team	**SCL-DeBERTa (Ours)**	0.955	**0.825**	0.829

This table presents the leaderboard results on the PAN 2025 benchmark. Our team, xxsu-team, with the proposed **SCL-DeBERTa** method, achieved top-tier performance across all three subtasks, ranking 3rd, 1st, and 2nd on Task1, Task2, and Task3, respectively. Compared to other strong baselines, our approach demonstrates superior robustness and adaptability in multi-author writing style analysis, establishing a new benchmark for this challenging task.

7. Conclusion And Future Work

7.1. Conclusion

In this study, we proposed *SCL-DeBERTa*, a novel framework that integrates supervised contrastive learning with the DeBERTa encoder to address the challenging task of multi-author writing style detection. By combining classification and contrastive losses, the model effectively enhances style differentiation and achieves precise detection of author transition boundaries.

7.2. Future Work

While *SCL-DeBERTa* has shown promising results, several avenues for future research remain open. First, we plan to explore the integration of additional pre-trained language models, such as GPT or LLAMA, to further enhance the model's stylistic representation capabilities. Second, extending the framework to handle multilingual datasets could broaden its applicability to diverse linguistic contexts. Third, we aim to investigate the impact of incorporating external knowledge, such as syntactic or semantic features, to improve the model's performance on the Hard subtask. Finally, optimizing the computational efficiency of the framework, particularly for large-scale datasets, will be a key focus to ensure its scalability in practical scenarios.

8. Acknowledgments

This work is supported by the National Social Science Foundation of China (24BYY080).

9. Declaration on Generative AI

During the preparation of this work, the authors utilized the DeepSeek language model and GPT-4.1 for grammaiand spelling checks. Following the use of this tool, the authors carefully reviewed and edited the contenas necessary and assume full responsibility for the content of this publication

References

[1] E. Zangerle, M. Mayerl, M. Potthast, B. Stein, Overview of the Multi-Author Writing Style Analysis Task at PAN 2025, in: G. Faggioli, N. Ferro, P. Rosso, D. Spina (Eds.), Working Notes of CLEF 2025 – Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings, CEUR-WS.org, 2025.

[2] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, D. Krishnan, Supervised contrastive learning, Advances in neural information processing systems 33 (2020) 18661–18673.

[3] P. He, X. Liu, J. Gao, W. Chen, Deberta: Decoding-enhanced bert with disentangled attention, arXiv preprint arXiv:2006.03654 (2020).

[4] A. Khan, M. Rai, K. Khan, S. Shah, F. Alvi, A. Samad, Team gladiators at pan: improving author identification: a comparative analysis of pre-trained transformers for multi-author classification, Working Notes of CLEF (2024).

[5] T. Lin, Y. Wu, L. Lee, Team nycu-nlp at pan 2024: integrating transformers with similarity adjustments for multi-author writing style analysis, Working Notes of CLEF (2024).

[6] H. Chen, Z. Han, Z. Li, Y. Han, A writing style embedding based on contrastive learning for multi-author writing style analysis., in: CLEF (Working Notes), 2023, pp. 2562–2567.

[7] J. Lv, Y. Yi, H. Qi, Team fosu-stu at pan: Supervised fine-tuning of large language models for multi author writing style analysis, Working Notes of CLEF (2024).

[8] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, W. Chen, et al., Lora: Low-rank adaptation of large language models., ICLR 1 (2022) 3.

[9] T. Chen, S. Kornblith, M. Norouzi, G. Hinton, A simple framework for contrastive learning of visual representations, in: International conference on machine learning, PmLR, 2020, pp. 1597–1607.