

CLEF 2025 PAN: Multi-Author Writing Style Analysis

Ali Muhammad Asad¹, Musaib², Syed Muhammad Areeb Kazmi³ and Sarim Tahir⁴

¹Department of Computer Science, Habib University

²Department of Computer Science, Habib University

³Department of Computer Science, Habib University

⁴Department of Computer Science, Habib University

Abstract

Multi-Author Writing Style Analysis is a crucial task in computational linguistics and authorship attribution, aimed at identifying points of transition within a document. This task has significant applications in forensic linguistics, plagiarism detection, and content verification. Over the years, various approaches have been explored, ranging from traditional stylometric techniques to advanced deep learning models. Early research relied on lexical and syntactic features to detect style changes, but these methods faced limitations in handling nuanced transitions. Recent advancements, particularly the integration of transformer-based models, have significantly improved the accuracy of style change detection. This paper aims to provide an overview of the tasks, the datasets, the evaluation metrics, and the baseline models for the Multi-Author Writing Style Analysis Lab of the PAN track at CLEF 2025.

1. Introduction

Writing style analysis is a fundamental problem in computational linguistics, with applications in authorship attribution, forensic text analysis, and collaborative content verification. A challenge in this domain is detecting when authorship changes within a multi-author document, a task known as style change detection. This problem has been the focus of multiple iterations of the PAN shared tasks, which have provided benchmark datasets and evaluation frameworks to advance the field.

Early approaches to style change detection leveraged handcrafted features such as n-grams, part-of-speech tag frequencies, and sentence structure analysis. These methods, while effective for basic segmentation tasks, struggled with complex, subtle changes in authorial style. The introduction of deep learning and pre-trained transformer models revolutionized the field, allowing for more nuanced analysis of writing styles. Models like BERT and DeBERTa have demonstrated superior performance in detecting stylistic shifts at both the paragraph and sentence levels.

This paper aims to explore existing methods and challenges in multi-author writing style analysis, reviewing state-of-the-art approaches such as deep learning models, transfer learning, and ensemble techniques. We will discuss the datasets, evaluation metrics, and baseline models used in the previous iterations of Multi-Author Writing Style Analysis Lab of the PAN track, ultimately attempting to improve the performance of style change detection.

2. Literature Review

Traditional approaches to multi-author writing style analysis mainly rely on lexical and syntactic feature based methods such as TF-IDF vectorization, character n-grams, and POS tag frequencies. These methods have demonstrated moderate performance in detecting style change at the paragraph level, but struggle with sentence-level changes due to limited contextual understanding. For instance, the PAN 2021 task utilized lexical similarity measures to detect authorship changes, achieving an F1 score of 0.73 on multi-author documents, indicating that feature-based methods were more effective for

✉ aa07190@st.habib.edu.pk (A. M. Asad); mu07694@st.habib.edu.pk (Musaib); mk07202@st.habib.edu.pk (S. M. A. Kazmi); st07112@st.habib.edu.pk (S. Tahir)



document-wide analysis but struggled with fine-grained author attribution [1]. PAN 2022 introduced structural features such as indentation and average sentence length, further refining style change detection by incorporating discourse-level analysis [2]. However, these methods were constrained by their reliance on handcrafted features and were susceptible to topic influence.

PAN 2023 and 2024 introduced more powerful techniques such as transformer based methods. The top-performing models in PAN 2023 leveraged pre-trained transformers like DeBERTaV3 and BERT for paragraph-level classification, achieving state-of-the-art F1 scores of 0.83 for the hardest dataset variant [3]. These models incorporated contrastive learning and fine-tuning on task-specific datasets, allowing them to distinguish between subtle stylistic variations more effectively. However, their performance degraded when topic variations were minimized, highlighting the challenge of isolating style from content.

Another approach integrated contrastive learning with DeBERTaV3 and knowledge distillation for efficiency [3]. This method allowed models to refine their representations of stylistic shifts while maintaining computational feasibility. Additionally, multi-stage fine-tuning techniques were introduced in PAN 2024, further improving performance by integrating both syntactic and semantic embeddings for style differentiation [4]. Hybrid approaches were also explored, combining traditional and deep learning methods. One study used a stacking ensemble of lexical and neural models, achieving an F1 score of 0.82 on medium-difficulty datasets [3]. Another team experimented with sentence-level style change detection using hierarchical clustering and transformer embeddings, demonstrating the potential for finer granularity in authorship attribution [4].

Performance evaluation has been standardized using the macro-averaged F1-score, with separate rankings for easy, medium, and hard datasets. PAN 2023 results highlighted the superiority of transformer-based models over traditional techniques, with the best-performing systems leveraging multi-stage fine-tuning, semantic similarity post-processing, and ensemble classifiers [4]. However, performance variations across dataset difficulty levels suggest that improvements in data augmentation and domain adaptation are needed.

References

- [1] E. Zangerle, M. Mayerl, M. Potthast, B. Stein, Overview of the style change detection task at pan 2021, CLEF 2021 - Conference and Labs of the Evaluation Forum, September 21-24, 2021, Bucharest, Romania (2021).
- [2] E. Zangerle, M. Mayerl, M. Potthast, B. Stein, Overview of the style change detection task at pan 2022, CLEF 2022 - Conference and Labs of the Evaluation Forum, September 5-8, 2022, Bologna, Italy (2022).
- [3] E. Zangerle, M. Mayerl, M. Potthast, B. Stein, Overview of the multi-author writing style analysis task at pan 2023, CLEF 2023 - Conference and Labs of the Evaluation Forum, September 18-21, 2023, Thessaloniki, Greece (2023).
- [4] E. Zangerle, M. Mayerl, M. Potthast, B. Stein, Overview of the multi-author writing style analysis task at pan 2024, CLEF 2024: Conference and Labs of the Evaluation Forum, September 09-12, 2024, Grenoble, France (2025).