

Team foshan-university-of-guangdong at PAN: Adaptive Entropy-Based Stability-Plasticity for Multi-Author Writing Style Analysis

Notebook for the PAN Lab at CLEF 2024

Xurong Liu^{1,*}, Hui Chen² and Jiajun Lv¹

¹ Foshan University, Foshan , Guangdong, China

² Shenzhen University, shenzhen , Guangdong, China

Abstract

In this paper, we address the Multi-Author Writing Style Analysis task for PAN 2024, which involves detecting style changes at paragraph levels within multi-author documents. To tackle this problem, we adopt the Entropy-based Stability-Plasticity (ESP) method, which dynamically adjusts the learning rates of different neural network layers based on their entropy values. This approach effectively balances stability and plasticity, allowing the model to retain essential knowledge from previous tasks while efficiently learning new information, thereby mitigating catastrophic forgetting. Our experiments, conducted on datasets of varying difficulty levels (Easy, Medium, Hard), demonstrate that ESP significantly outperforms traditional methods in detecting writing style changes. The results highlight the effectiveness of ESP in leveraging prior knowledge and reducing interference between tasks, making it a robust framework for continuous learning in text analysis applications.

Keywords

style change detection, ESP, lifelong learning

1. Introduction

The style change detection task aims to identify text positions within a given multi-author document at which the author switches[1]. When there are no comparison texts are provided, if multiple authors together have written a text, the only way that find evidence for this fact is style change detection to detect plagiarism in a document. Likewise, style change detection can help to uncover gift authorships, to verify a claimed authorship, or to develop new technology for writing support[2]. Style change detection is a branch of authorship verification focussing on the examination of a document for the different authorial style[3]. The application areas of writing style change detection range from plagiarism detection, cyber security, and forensics and currently in fake news detection [4, 5, 6]. Endeavors to detect changes in writing styles have been done under author dimerization or clustering and style change detection [7, 8, 9].

The ability to continuously learn remains elusive for deep learning models, which cannot accumulate knowledge in their weights when learning new tasks. For the task Multi-Author Writing Style Analysis 2024 of PAN that has three different difficulty datasets: Easy, Medium, and Hard, we adopted a method called Entropy-based Stability-Plasticity (ESP) [10] for lifelong learning to address this problem, which can decide dynamically how much each model layer should be modified via a plasticity factor. The results show the robust framework of the approach in leveraging prior knowledge by reducing interference, offering slight improvements over the baseline in maintaining performance across sequential tasks.

2. Background

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

*Corresponding author.

✉ Liu_xurongx@163.com (X. Liu); chenhui_yeah0101@163.com (H. Chen); lvjiajun.96@gmail.com (J. Lv)

🆔 0009-0000-4386-5336 (X. Liu); 0009-0007-2695-3220 (H. Chen); 0000-0002-8755-5310 (J. Lv)

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



For the task of Multi-Author Writing Style Analysis, some studies employ the use of feature combinations such as lexical, syntactic and character features to analyze the variance in the styles of writing by different authors [11, 12, 13, 14]. Some of these methods rely on the analysis of the different stylometric features to detect stylistic changes in a document [15], while others adapted the outlier detection methods used in plagiarism detection problems. In addition some studies investigated the use of artificial neural networks to solve this problem [16, 15].

The task Multi-Author Writing Style Analysis 2024 of PAN is defined as, for a given text, finding all positions of writing style change on the paragraph level (i.e., for each pair of consecutive paragraphs, assess whether there was a style change) [1]. The simultaneous shift in authorship and topic will be carefully controlled; Task 1, Task 2, and Task 3 correspond to datasets of three different difficulty levels: Easy, Medium, and Hard. The corresponding dataset descriptions are as follows:

- **Easy:** Easy: The paragraphs of a document cover a variety of topics, allowing approaches to use topic information to detect authorship changes.
- **Medium:** The topical variety in a document is small (though still present), forcing the approaches to focus more on style to effectively solve the detection task.
- **Hard:** All paragraphs in a document are on the same topic.

Artificial neural networks learn in a bounded environment, where the input distribution is assumed to be fixed. When the input distribution changes, the model must adapt its weights to perform correctly on the new task. Due to those modifications, the model overwrites previously learned patterns, creating interference between old and new tasks, causing a problem known as catastrophic forgetting [17, 18]. The lifelong learning methods to alleviate catastrophic forgetting can be categorized into three classes, based on how storing and using the task-specific information throughout the sequential learning process, the replay methods, the regularization-based methods, and the parameter isolation methods [19]. Vladimir Araujo proposes the ESP method, which relies on an entropy-based criterion to decide how much a model has to modify the weights in each of its layers, which performs well compared to the Stability version [10], and outperforms all baselines when trained on all experiments such as Replay.

3. System Overview

The Entropy-Based Stability-Plasticity (ESP) [10] method for lifelong learning utilizes an entropy-based criterion to manage the trade-off between stability and plasticity. The plasticity factor is a crucial component in ESP, determining how much each layer of the neural network should be updated. This factor is computed using entropy to assess the importance of the parameters in each layer. The plasticity factor P_i for the i -th layer can be expressed as:

$$P_i = \frac{H_i}{\sum_{j=1}^L H_j} \quad (1)$$

where H_i is the entropy of the i -th layer and L is the total number of layers.

The entropy H_i for the i -th layer is calculated based on the activations of the neurons in that layer. The entropy helps in identifying how much information is being processed by the layer. The entropy H_i for the i -th layer can be defined as:

$$H_i = - \sum_{k=1}^{N_i} p_k \log p_k \quad (2)$$

where N_i is the number of neurons in the i -th layer and p_k is the probability associated with the k -th neuron's activation.

ESP uses the plasticity factor to scale the gradients during backpropagation. This ensures that layers with higher entropy (more important) receive smaller updates, preserving their stability, while layers

with lower entropy (less important) are updated more, enhancing plasticity. The gradient update for the i -th layer is scaled by its plasticity factor P_i :

$$\Delta\theta_i = P_i \cdot \nabla\theta_i \quad (3)$$

where $\nabla\theta_i$ is the gradient of the loss with respect to the parameters of the i -th layer. The ESP training process involves the following steps:

1. Forward Pass: Compute the output and activations of each layer.
2. Entropy Calculation: Calculate the entropy for each layer.
3. Plasticity Factor Calculation: Determine the plasticity factors based on the entropies.
4. Backward Pass: Scale the gradients using the plasticity factors and update the model parameters.

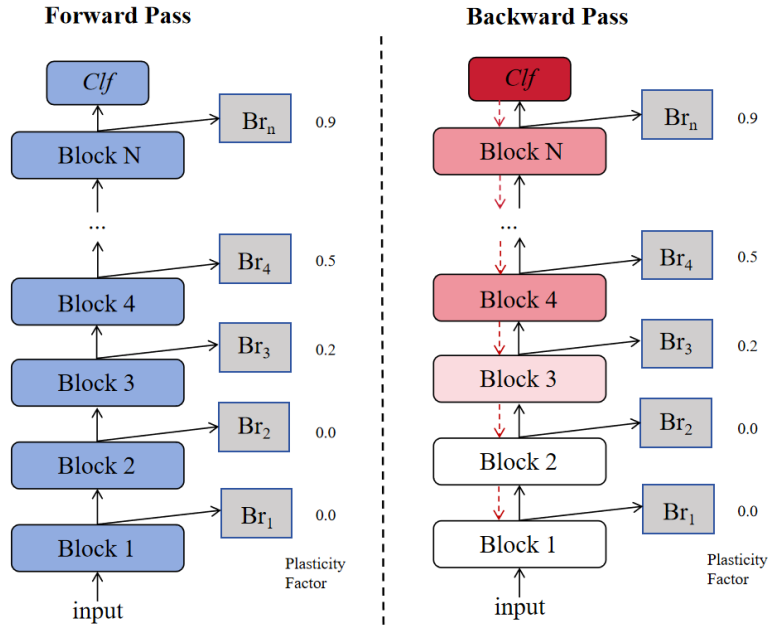


Figure 1: Overview of the method. During the forward step, the backbone processes an example and generates prediction and plasticity factor values for each block (left). During the backward pass, the plasticity factor is used to adjust the final amount of modification each layer will have (right) [10].

ESP effectively balances stability and plasticity by dynamically adjusting the learning rate for different layers based on their entropy. This method ensures that important layers (with high entropy) are preserved while less important layers (with low entropy) are more flexible to learn new tasks. These formulas and the underlying methodology provide a robust framework for lifelong learning, addressing the challenge of catastrophic forgetting while allowing the model to adapt to new information.

We use BERT [20] as the encoder. For the decoder, following the original BERT model, we use the first token (special token [CLS]) of the sequence and a classifier to predict the class. Additionally, we use the default BERT vocabulary in our experiments. We utilize the Adam optimizer with a learning rate of 3×10^{-5} and a training batch size of 32. To enhance the training for Task 2 and Task 3, we adjusted the training sequence of the data to hard \rightarrow medium \rightarrow easy.

4. Results

Following the above experiment design, the results are as table 1 follows:

Table 1

Overview of the F1 accuracy for the multi-author writing style task in detecting at which positions the author changes for task 1, tas 2, and task 3.

Approach	Task 1	Task 2	Task 3
emerald-callable	0.517	0.394	0.352
Baseline Predict 1	0.466	0.343	0.320
Baseline Predict 0	0.112	0.323	0.346

5. Conclusion

In this work, we addressed the Multi-Author Writing Style Analysis task for PAN 2024, which involves detecting style changes at paragraph levels in multi-author documents. The approach, Entropy-based Stability-Plasticity (ESP), effectively manages the trade-off between stability and plasticity in lifelong learning scenarios by dynamically adjusting the learning rates of different network layers based on their entropy values. This ensures that layers with high entropy, which are deemed more important, receive smaller updates to preserve stability, while layers with low entropy, considered less critical, are updated more to enhance plasticity. Our experiments utilized BERT as the encoder and demonstrated the effectiveness of ESP across different difficulty levels of datasets (Easy, Medium, Hard). The results showed that ESP outperforms traditional methods by effectively leveraging prior knowledge and reducing interference between tasks. Specifically, ESP’s adaptive gradient scaling mechanism allows the model to retain essential information from previous tasks while efficiently learning new tasks, thus mitigating the issue of catastrophic forgetting.

In conclusion, the ESP method provides a robust framework for continuous learning in multi-author writing style analysis, offering slight improvements over the baseline in maintaining performance across sequential tasks. Future work may explore the integration of ESP with other neural architectures and its application to additional domains beyond text analysis.

References

- [1] 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.
- [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] V. A. Oloo, C. Otieno, L. A. Wanzare, A literature survey on writing style change detection based on machine learning: State-of-the-art–review, *Int. J. Comput. Trends Technol.* 70 (2022) 15–32.
- [4] D. Castro-Castro, C. A. Rodríguez-Lozada, R. Muñoz, Mixed style feature representation and b-maximal clustering for style change detection., in: *CLEF (Working Notes)*, 2020.
- [5] A. Iyer, S. Vosoughi, Style change detection using bert., *CLEF (Working Notes)* 93 (2020) 106.

- [6] C. Zuo, Y. Zhao, R. Banerjee, Style change detection with feed-forward neural networks., CLEF (Working Notes) 93 (2019).
- [7] S. Nath, Style change detection using siamese neural networks., in: CLEF (Working Notes), 2021, pp. 2073–2082.
- [8] P. Rosso, F. Rangel, M. Potthast, E. Stamatatos, M. Tschuggnall, B. Stein, Overview of pan'16: new challenges for authorship analysis: cross-genre profiling, clustering, diarization, and obfuscation, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction: 7th International Conference of the CLEF Association, CLEF 2016, Évora, Portugal, September 5-8, 2016, Proceedings 7, Springer, 2016, pp. 332–350.
- [9] E. Zangerle, M. Mayerl, M. Potthast, B. Stein, Overview of the style change detection task at pan 2020., CLEF (Working Notes) 93 (2020).
- [10] V. Araujo, J. Hurtado, A. Soto, M.-F. Moens, Entropy-based stability-plasticity for lifelong learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 3721–3728.
- [11] 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, 2021.
- [12] M. L. Brocardo, I. Traore, S. Saad, I. Woungang, Authorship verification for short messages using stylometry, in: 2013 International Conference on Computer, Information and Telecommunication Systems (CITS), IEEE, 2013, pp. 1–6.
- [13] M. P. Kuznetsov, A. Motrenko, R. Kuznetsova, V. V. Strijov, Methods for intrinsic plagiarism detection and author diarization., in: CLEF (Working notes), 2016, pp. 912–919.
- [14] K. Safin, R. Kuznetsova, Style breach detection with neural sentence embeddings., in: CLEF (Working Notes), 2017.
- [15] A. Rexha, M. Kröll, H. Ziak, R. Kern, Authorship identification of documents with high content similarity, *Scientometrics* 115 (2018) 223–237.
- [16] M. Kestemont, M. Tschuggnall, E. Stamatatos, W. Daelemans, G. Specht, B. Stein, M. Potthast, Overview of the author identification task at pan-2018: cross-domain authorship attribution and style change detection, in: Working Notes Papers of the CLEF 2018 Evaluation Labs. Avignon, France, September 10-14, 2018/Cappellato, Linda [edit.]; et al., 2018, pp. 1–25.
- [17] M. McCloskey, N. J. Cohen, Catastrophic interference in connectionist networks: The sequential learning problem, in: *Psychology of learning and motivation*, volume 24, Elsevier, 1989, pp. 109–165.
- [18] R. Ratcliff, Connectionist models of recognition memory: constraints imposed by learning and forgetting functions., *Psychological review* 97 (1990) 285.
- [19] M. De Lange, R. Aljundi, M. Masana, S. Parisot, X. Jia, A. Leonardis, G. Slabaugh, T. Tuytelaars, A continual learning survey: Defying forgetting in classification tasks, *IEEE transactions on pattern analysis and machine intelligence* 44 (2021) 3366–3385.
- [20] 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).