iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IX **Month of publication:** September 2024

DOI: https://doi.org/10.22214/ijraset.2024.64168

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IX Sep 2024- Available at www.ijraset.com

Deep Learning for Stylometry and Authorship Attribution: a Review of Literature

Nishchal Sharma[1], Ajay Kumar[2]
[1]*Department of Computer Science, Govt. College Kullu*
[2]*Department of English, Govt. College Kullu*

Abstract: ***The application of deep learning techniques to stylometry and authorship attribution has emerged as a promising frontier in computational linguistics, offering new possibilities for understanding literary style and authorship in both historical and contemporary contexts. This review paper synthesizes recent advances in the use of deep learning models, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and transformer architectures, for identifying and attributing authorship based on stylistic analysis. We examine the effectiveness of these models in comparison to traditional statistical methods, highlighting their ability to capture complex linguistic patterns and nuances that are often overlooked by conventional approaches. Furthermore, we explore how deep learning models handle challenges such as multilingual texts, limited data, and variations across genres and periods. This review also addresses the interpretability of neural networks in the context of stylometry and discusses the implications of these methods for fields ranging from literary studies to digital forensics. By providing a comprehensive overview of the current state of research, this paper identifies key trends, challenges, and future directions for the application of deep learning to stylometry and authorship attribution.***

Keywords: Deep Learning, Stylometry, Authorship Attribution, Computational Linguistics, Neural Network

I. INTRODUCTION

The study of stylometry—analyzing the unique linguistic style of a text to attribute authorship—has long been a significant area of inquiry in literary studies, forensic linguistics, and digital humanities. Traditionally, authorship attribution has relied on statistical techniques that use handcrafted features, such as word frequency counts, sentence length, and syntactic patterns, to distinguish between authors. While these methods have achieved considerable success, they are often limited by their reliance on predefined features and their inability to capture the more subtle and complex aspects of an author's style. With the advent of deep learning, a new paradigm for stylometric analysis has emerged, offering the potential to move beyond these limitations.

Deep learning, a subset of machine learning that leverages neural networks to learn from large datasets, has transformed various fields, from computer vision to natural language processing. Its ability to automatically learn hierarchical representations from data makes it particularly well-suited for the task of authorship attribution, where capturing intricate patterns in text is crucial. Recent studies have shown that deep learning models, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and transformers, can significantly outperform traditional methods in identifying authorship, even in challenging scenarios involving multiple authors, short texts, and cross-genre analysis.

This paper provides a comprehensive review of the literature on the application of deep learning to stylometry and authorship attribution. We examine various neural network architectures and their performance compared to conventional methods, highlighting their advantages in capturing complex stylistic features. Additionally, we explore the challenges faced by these models, including interpretability, data limitations, and generalizability across different languages and genres. By synthesizing the current state of research, this review aims to illuminate the potential of deep learning for advancing the field of stylometry and identify areas for future exploration.

II. PREVIOUS WORK

The paper [1] investigates the effectiveness of stylometric (writing style) and emotion-based features for detecting hate speech across three languages: English, Slovene, and Dutch. The authors conduct experiments in both in-domain and cross-domain setups to assess whether these features can robustly classify textual content as hate or non-hate speech. The study models two linguistic phenomena: function word usage (to capture writing style) and emotion expression in hateful messages. Results indicate that stylometric and emotion-based features are strong indicators of hate speech, consistently performing well across different domains and languages. When combined, these features outperform traditional word and character n-gram features, especially in cross-domain settings.

©IJRASET: All Rights are Reserved

Furthermore, integrating these features with deep learning models through a majority-voting ensemble significantly improves detection performance. The study concludes that the unique stylistic and emotional characteristics of hate speech make these features valuable for enhancing cross-domain robustness in hate speech detection tasks.

The effectiveness of parallel stylometric document embeddings for authorship attribution by applying a novel approach to literary texts in seven different languages was studied in this paper [2] . The study analyzes 7,051 unique 10,000-token chunks derived from 700 documents annotated with Part-of-Speech (PoS) tags and lemmas. Four document embedding models were created using the Stylo R package (based on words, lemmas, PoS trigrams, and PoS masks) and one model using mBERT for each language. Various combinations of these embeddings were derived (average, product, minimum, maximum, and L2 norm), both including and excluding mBERT-based embeddings. Several perceptrons were trained on portions of the dataset to determine optimal weights for a weighted combination approach. The study compared standalone and composite embeddings for classification accuracy, precision, recall, and F1-score. The results indicate that most composite methods outperform the baselines for each language, with some methods surpassing all baselines across all languages. Additionally, the inclusion of mBERT embeddings did not significantly improve performance.

The need for effective authentication systems in online examinations, particularly to address the limitations of traditional plagiarism checkers. Traditional plagiarism detection tools are primarily designed to identify verbatim copying of text but struggle with non-verbatim plagiarism, such as text rewritten in different words or translated, and text generated by text-converters. This makes them less effective for verifying the authenticity of online examination submissions. To overcome these limitations, the paper [3] proposes a stylometric analysis approach, which involves analyzing and matching the writing patterns of the submitted text to those of the individual author. By extracting and analyzing 27 stylistic features specific to each writer, this method aims to verify if a document was authored by a particular individual rather than checking for copied content. This approach helps prevent the use of text-converters or translators to disguise the original source of the text. Additionally, the paper includes a comparative analysis of simpler algorithms for stylometric verification, concluding that Artificial Neural Networks provide the most accurate and precise solution for verifying authorship in online examinations.

The paper [4] explores the problem of authorship identification, which involves determining the author of an anonymous or unknown document by analyzing various text features. This task is crucial in Natural Language Processing and has diverse applications such as identifying anonymous authors, aiding crime investigations, detecting plagiarism, and uncovering ghostwriters. The study moves beyond traditional methods that rely on character n-grams by employing advanced feature engineering and different text-based models. It investigates a range of stylometric features and identifies those that significantly improve model performance. The proposed methodology is tested on a subset of the Reuters news corpus, which includes texts from 50 different authors on the same topic. Results show that using document fingerprinting features enhances classifier accuracy, and Principal Component Analysis (PCA) further improves these results. The paper also compares the proposed approach with existing methods in the field, demonstrating its effectiveness in authorship identification.

The paper [5] discusses the authors' involvement in the Bots and Gender Profiling task at PAN 2019, which focuses on identifying whether a profile is created by a bot or a human and, if human, classifying the gender of the profile. The approach used by the authors relies on 27 language-independent stylometry features, including 18 character-based and 9 emotion-based features. For the English language, their system achieved high accuracy on the training dataset, with scores of 0.97 for distinguishing bots from humans and 0.80 for gender classification. In Spanish, the accuracies were 0.93 for bot vs. human and 0.75 for gender classification. On test dataset 1, the English results were 0.92 for bot vs. human and 0.76 for gender classification, while Spanish results were 0.86 and 0.75, respectively. In test dataset 2, the system maintained strong performance with accuracy scores of 0.92 and 0.76 for English and 0.88 and 0.72 for Spanish, respectively.

Authorship analysis (AA) involves uncovering the hidden characteristics of authors through textual data, focusing on their identity and sociolinguistic traits as reflected in their writing style. This analysis is crucial for fields like cybercrime investigation, psycholinguistics, and political socialization. Traditionally, AA techniques have relied heavily on manual feature engineering, making them dependent on the specific scenario or dataset.

The paper [6] introduces a neural network-based approach to improve authorship analysis by incorporating various linguistic features into the distributed representation of words. This method aims to replicate human sentence composition processes by learning writing style representations from unlabeled texts. The proposed models extract topical, lexical, syntactical, and character-level feature vectors as stylometrics. The approach is evaluated across different tasks—authorship characterization, identification, and verification—using datasets from Twitter, blogs, reviews, novels, and essays.

The results indicate that this method significantly outperforms traditional static stylometrics, dynamic n-grams, Latent Dirichlet Allocation, Latent Semantic Analysis, and other baseline techniques, including word2vec and paragraph vector representations.

A thesis [7] investigates the pair-wise impostor finding problem, which involves determining if two user accounts on social networks, potentially based on their short conversational texts, are controlled by the same author. It evaluates two approaches to solving this problem:

One approach uses stylometric authorship attribution methods combined with the Doppelgänger Finder. This method involves three stylometric techniques—Cosine Delta, SVM, and CNN—along with various feature sets. The CNN model consistently performs the best across all datasets, with character 3-grams and word 1-grams being the most effective features. The second approach is a novel network analysis-based method, tested on the Opsahl Facebook-like Social Network dataset with added Tweets from the Sentiment140 Twitter dataset. This method combines network features with stylometric features. While neither network nor stylometric features alone yield significant results, their combination achieves a weighted F1-score of 0.7. When comparing the two methods, the SVM model combined with the Doppelgänger Finder achieves a pair-wise score of no more than 0.5, which is less effective than the 0.7 score attained by the network analysis-based method.

The paper [8] explores the use of stylometry as an authentication method for verifying authorship, leveraging the unique and non-fabricable nature of individual writing styles. It evaluates how effectively stylometry can distinguish between writings on the same topic from different users. The study achieved a 74% accuracy in identifying the actual authors and suggests that incorporating additional features could increase accuracy to over 90%. The paper also establishes a threshold for user authentication and examines how combining textual features can enhance authenticity verification. Additionally, it assesses the impact of data cleaning methods, such as removing stop words and punctuation, on the overall detection results.

The paper [9] discusses recent advancements in artificial neural networks and deep learning, particularly their application to source code analysis. While deep learning has excelled in fields like image, video, and speech processing, it has struggled with capturing the structural and behavioral aspects of source code, often relying on manual feature engineering for tasks such as author identification and code quality analysis. To address this, the paper introduces SnapCode, a novel approach that processes code snapshots instead of individual tokens. SnapCode employs a deep convolutional neural network and transfer learning to extract structural features from the source code. The study demonstrates that while simple networks fail to capture these features, SnapCode's deep network and transfer learning approach yields superior results. The approach is validated through the task of author detection, or "code stylometry," which benefits from SnapCode's ability to learn complex features and capture behavioral aspects of the code. The study [10] compares various methods for feature extraction in text representation for gender and author recognition tasks. It evaluates the traditional Bag-of-Words (BoW) approach, which uses frequency vectors of features like lemmas, word forms, bigrams of grammatical classes, and punctuation. It also examines the fastText algorithm for vector-based text representation. The methods are tested on two Polish literary text collections from the late 19th and early 20th centuries: one with 99 novels from 33 authors and another with 888 novels from 58 authors. The results indicate that the effectiveness of feature types varies depending on the corpus, with grammatical bigrams and semantic features (such as the most frequent 1000 lemmas) showing the best performance. The study also highlights the importance of properly splitting corpora into training and testing sets for accurate results. Recent advancements in neural language models (LMs) have raised concerns about their potential misuse in spreading misinformation. This work [11] investigates the effectiveness of stylometry, a technique used to detect machine-generated fake news by analyzing stylistic differences between human-written and machine-generated texts. While stylometry has been successful in source attribution and detecting misinformation in human texts, it struggles with machine-generated content. LMs produce stylistically consistent text regardless of intent, making it difficult for stylometric methods to distinguish between legitimate and deceptive uses. The study introduces two benchmarks that demonstrate the stylistic similarity between malicious and legitimate uses of LMs, such as in auto-completion and editing-assistance. The findings suggest that while stylometry can prevent impersonation, it is inadequate for detecting machine-generated misinformation. The paper calls for the development of more effective benchmarks and non-stylometric methods for detecting false information, and emphasizes the need for interdisciplinary approaches involving NLP, social networks, information security, and human-computer interaction.

The rise of large language models (LLMs) that generate realistic text and images has raised ethical concerns, prompting research into distinguishing AI-generated content from human-authored material. This study [12] introduces **StyloAI**, a data-driven model designed to identify AI-generated texts using 31 stylometric features and a Random Forest classifier. Tested on two multi-domain datasets, StyloAI achieved accuracy rates of 81% and 98% on the AuTextification and Education datasets, respectively. This model outperforms existing state-of-the-art methods and sheds light on the distinctive characteristics of AI-generated versus human-authored texts.

III.CONCLUSION

In conclusion, stylometry has proven to be a powerful tool for analyzing and understanding textual content by revealing unique characteristics of an author's writing style. Over the years, stylometric methods have been instrumental in a range of applications, including authorship attribution, verification, and profiling. The ability to quantify and analyze stylistic features provides valuable insights into both individual and collective writing behaviors, aiding in tasks from detecting plagiarism to identifying the authorship of anonymous texts. Recent advancements in stylometric techniques, particularly those incorporating machine learning and deep learning approaches, have enhanced the accuracy and applicability of stylometric analysis. Despite these advancements, challenges remain, especially in adapting stylometric methods to the rapidly evolving landscape of digital content and AI-generated text. The effectiveness of stylometry in these new contexts will depend on ongoing innovation and refinement of methods. As we move forward, it is essential to continue exploring the potential of stylometry while addressing its limitations. This includes developing more sophisticated models that can handle diverse and complex text types, as well as integrating stylometric analysis with other technological and methodological approaches. By doing so, stylometry can remain a relevant and powerful tool in the ever-changing world of textual analysis, contributing to our understanding of language, authorship, and digital content.

REFERENCES

[1] Markov, N. Ljubešić, D. Fišer, and W. Daelemans, "Exploring Stylometric and Emotion-Based Features for Multilingual Cross-Domain Hate Speech Detection," in Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, O. De Clercq, A. Balahur, J. Sedoc, V. Barriere, S. Tafreshi, S. Buechel, and V. Hoste, Eds., Online: Association for Computational Linguistics, Apr. 2021, pp. 149–159. Accessed: Sep. 05, 2024. [Online]. Available: https://aclanthology.org/2021.wassa-1.16

[2] M. Škorić, R. Stanković, M. Ikonić Nešić, J. Byszuk, and M. Eder, "Parallel Stylometric Document Embeddings with Deep Learning Based Language Models in Literary Authorship Attribution," Mathematics, vol. 10, no. 5, Art. no. 5, Jan. 2022, doi: 10.3390/math10050838.

[3] A. I. Khan, S. Jain, P. Sharma, V. Deep, and D. Mehrotra, "Stylometric Analysis of Writing Patterns Using Artificial Neural Networks," in 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), Sep. 2021, pp. 29–35. doi: 10.1109/3ICT53449.2021.9582095.

[4] S. Yadav, S. S. Rathore, and S. S. Chouhan, "Authorship Identification Using Stylometry and Document Fingerprinting," in Big Data Analytics, L. Bellatreche, V. Goyal, H. Fujita, A. Mondal, and P. K. Reddy, Eds., Cham: Springer International Publishing, 2020, pp. 278–288. doi: 10.1007/978-3-030-66665-1_18.

[5] S. Ashraf, O. Javed, M. Adeel, H. Ali, and R. M. A. Nawab, "Bots and Gender Prediction Using Language Independent Stylometry-Based Approach".

[6] S. H. H. Ding, B. C. M. Fung, F. Iqbal, and W. K. Cheung, "Learning Stylometric Representations for Authorship Analysis," IEEE Trans. Cybern., vol. 49, no. 1, pp. 107–121, Jan. 2019, doi: 10.1109/TCYB.2017.2766189.

[7] L. a. Y. Maagendans, "Impostor Finding Using Stylometry and Network Analysis," Master Thesis, 2021. Accessed: Sep. 05, 2024. [Online]. Available: https://studenttheses.uu.nl/handle/20.500.12932/38636

[8] K. Surendran, O. P. Harilal, P. Hrudya, P. Poornachandran, and N. K. Suchetha, "Stylometry Detection Using Deep Learning," in Computational Intelligence in Data Mining, H. S. Behera and D. P. Mohapatra, Eds., Singapore: Springer, 2017, pp. 749–757. doi: 10.1007/978-981-10-3874-7_71.

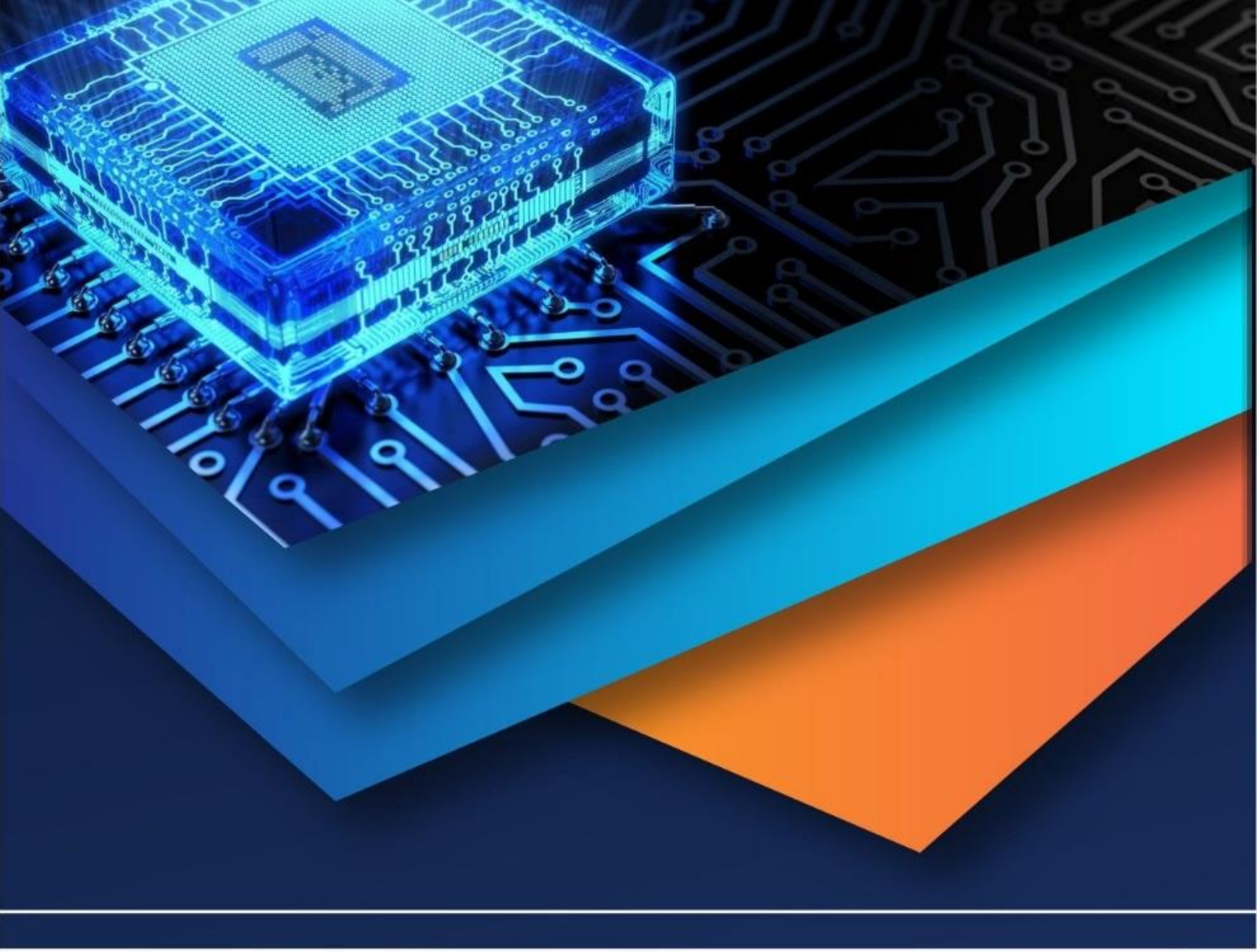
[9] "SnapCode - A Snapshot Based Approach to Code Stylometry | IEEE Conference Publication | IEEE Xplore." Accessed: Sep. 05, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9031980

[10] "Stylometry Analysis of Literary Texts in Polish | SpringerLink." Accessed: Sep. 05, 2024. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-91262-2_68

[11] T. Schuster, R. Schuster, D. J. Shah, and R. Barzilay, "The Limitations of Stylometry for Detecting Machine-Generated Fake News," Comput. Linguist., vol. 46, no. 2, pp. 499–510, Jun. 2020, doi: 10.1162/coli_a_00380.

[12] C. Opara, "StyloAI: Distinguishing AI-Generated Content with Stylometric Analysis," in Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky, A. M. Olney, I.-A. Chounta, Z. Liu, O. C. Santos, and I. I. Bittencourt, Eds., Cham: Springer Nature Switzerland, 2024, pp. 105–114. doi: 10.1007/978-3-031-64312-5_13.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)