

An MSc project at the Netherlands Forensic Institute

Deep learning for author recognition

MSc project

This is a nice project for a MSc. student specializing in Deep Learning or likewise. The challenge for the student is to

- Apply and compare authorship attribution techniques based on DL (e.g., BERT) for authorship verification task (e.g., on transcriptions).
- Given that most of those algorithms have been used for language understanding and topic detection, how are their performance affected by the topic of the conversation?
- Can we use a topic masking technique in combination with such algorithms to fine-tune it for an authorship verification task?

References

- [1] Barlas, Georgios, and Efstathios Stamatatos. 2020. "Cross-Domain Authorship Attribution Using Pre-Trained Language Models." In *Artificial Intelligence Applications and Innovations*, 255–66. Springer International Publishing.
- [2] Sari, Yunita, Mark Stevenson, and Andreas Vlachos. 2018. "Topic or Style? Exploring the Most Useful Features for Authorship Attribution." In *Proceedings of the 27th International Conference on Computational Linguistics*, 343–53. Santa Fe, New Mexico, USA: Association for Computational Linguistics.
- [3] Stamatatos, Efstathios. 2018. "Masking Topic-Related Information to Enhance Authorship Attribution." *Journal of the Association for Information Science and Technology* 69 (3): 461–73.