Deep learning approach to monitoring misinformation and hate speech on social media

Supervisors

dr hab. prof. UW Piotr Sankowski (IDEAS NCBR)

dr Tomasz Michalak (IDEAS NCBR)

Description

The increasing prevalence of social media has presented significant challenges in combating misinformation and hate speech. To address this issue, this doctoral project adopts a multidisciplinary approach, aiming to develop advanced deep learning methods specialized in effectively detecting and monitoring harmful content on social media platforms [1].

The primary objective of this research is to create and implement deep neural networks that can analyze and classify social media posts to identify instances of misinformation and hate speech. To achieve this, the project will explore natural language processing (NLP) techniques, allowing the models to extract meaningful information from textual content and comprehend context, thus enhancing the accuracy of harmful content identification. Furthermore, the research will focus on developing strategies to handle the dynamic and continuously evolving nature of data on social media platforms [2].

The potential misuse of generative models for generating deceptive or misleading content is another aspect that requires careful consideration. Therefore ensuring the ethical use of generative models in social media content analysis will also be a key aspect of this research.

The project acknowledges the significance of a comprehensive data analysis framework, and it will emphasize rigorous data preprocessing and validation methodologies. Utilizing large-scale datasets from diverse social media platforms, the models will be trained and evaluated to ensure their effectiveness in real-world scenarios [3].

The ultimate goal of this endeavor is to contribute to the development of an efficient and reliable deep learning-based monitoring system for social media content. Such a system will have far-reaching implications for societal well-being by safeguarding users from harmful and misleading content, thus fostering a safer and more inclusive online environment. Through a multidisciplinary approach encompassing deep learning and data analysis, this research seeks to advance technology-driven solutions in combating misinformation and promoting a more responsible use of social media.

Requirements

- Good knowledge of deep learning, including practical experience with programming in Python and relevant libraries (PyTorch, TensorFlow, Keras)
- Understanding of social media platforms
- Data collection and annotation skills
- A strong understanding of ethical concerns related to monitoring social media content, user privacy, and data protection
- Strong analytical and problem-solving skills
- Creativity and innovation
- Collaboration and communication.

References

[1] Del Vicario, M., Zollo, F., Caldarelli, G., Scala, A., & Quattrociocchi, W. (2017). Mapping social dynamics on Facebook: The Brexit debate. Social Networks, 50, 6-16.

[2] Zervopoulos, A., Alvanou, A. G., Bezas, K., Papamichail, A., Maragoudakis, M., & Kermanidis, K. (2020). Hong Kong protests: using natural language processing for fake news detection on twitter. In Artificial Intelligence Applications and Innovations: 16th IFIP WG 12.5 International Conference, AIAI 2020, Neos Marmaras, Greece, June 5–7, 2020, Proceedings, Part II 16 (pp. 408-419). Springer International Publishing.

[3] Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, *35*(8), 1798-1828.