The Case for Fake News Detection as a Multimodal Deep Learning Task

Steven van de Graaf Michal Nauman Azamat Omuraliev

Introduction

When it comes to Fake News Detection (FND), generally there are 4 strategies, which focus on knowledge, credibility, style and propogation (Zhou et al., 2019).

This poster concerns itself with the latter 3 strategies. To this end, we have developed and implemented a Multimodal Deep Learning model, consisting of a Graph Convolutional Network (GCN) (Kipf & Welling, 2017) which employs the propogation strategy, and a Hierarchical Attention Network (HAN) (Yang et al. 2016), which employs the credibility and style strategies.

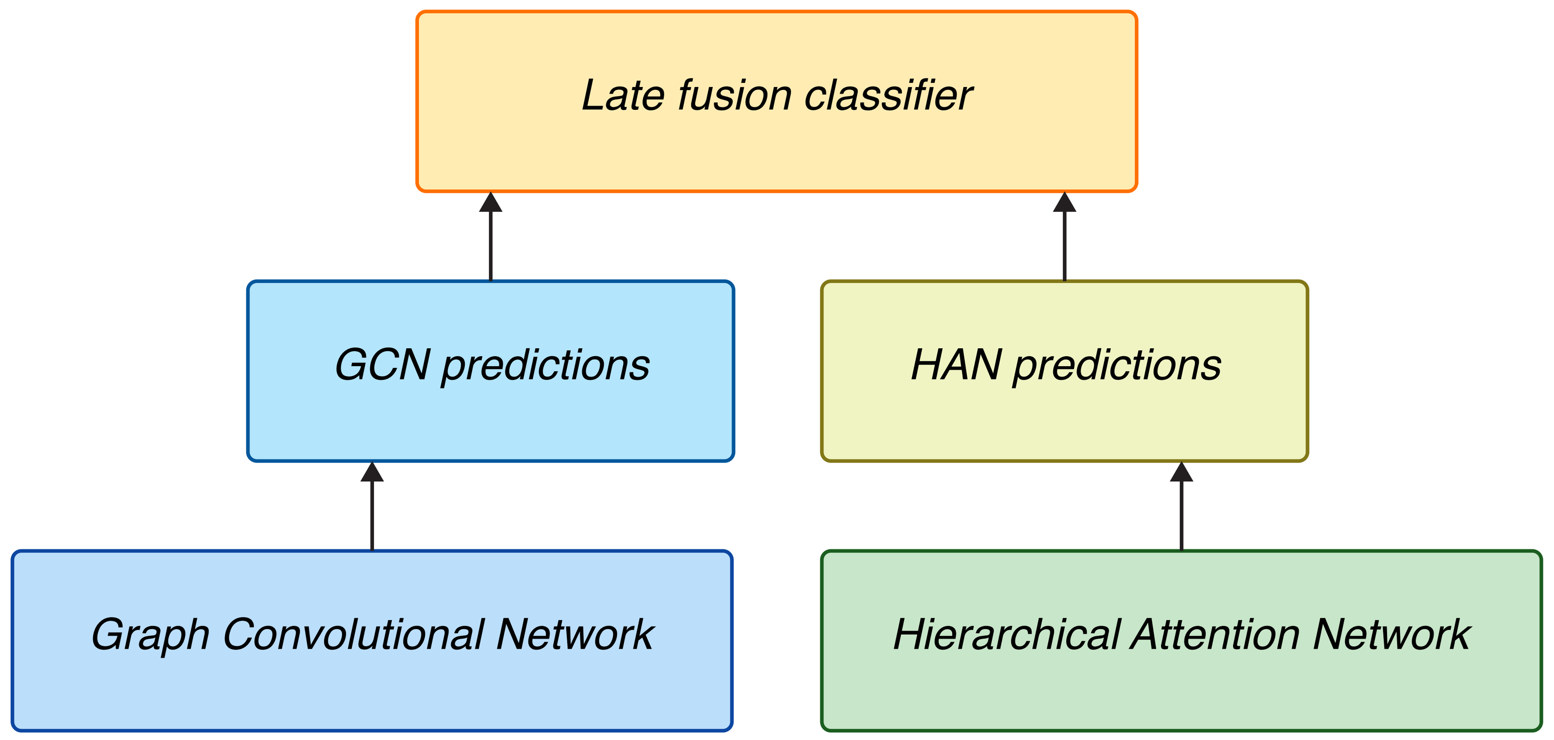

Figure 1. Our multimodal model architecture

Model

Graph Convolutional Network (GCN)

- The adjacency matrix consists of twitter users and articles
- The features are the Bag-of-Words (BoW) representation of the first paragraph of the article

Hierarchical Attention Network (HAN)

- Uses a bi-directional LSTM + Attention for both the attentative sentence encoder, as well as the attentative document encoder
- Uses concat. of GloVe and ELMo word embeddings
- Uses 100-dimensional sentence and document embeddings

Late fusion classifier

- Weighs the binary class probabilities of the GCN and HAN models evenly (50-50)

Experiments

Dataset

- The FakeNewsNet dataset was used, as published by Shu et al. (2018), consisting of 422 articles with their respective social context (tweets, follows, all fully anonimized)
- There are 211 fake and 211 real articles (balanced)
- There are 240 articles which come from PolitiFact. The remaining 182 articles come from BuzzFeed.

Training procedure

- The GCN and HAN models were trained seperately. The late fusion classifier was not trained.
- The GCN model was trained using default parameters (200 epochs, learning rate 0.01, hidden dim. 16, dropout probability 0.5)
- The HAN model was pre-trained on a larger Fake News dataset containing over 20k articles, before being fine-tuned by means of transfer learning for 20 epochs, using a learning rate of 0.001 and batch size of 2.

Baselines

- Both unimodal models (GCN and HAN) serve as baselines for our Multimodal model, as they can be directly compared

Results

- Due to the potentially high cost of missclasification, two metrics were used: classification accuracy and Area Under Curve (AUC) (the area under the ROC-curve)
- Inspection of the AUC metric revealed that for all models, the distance from the decision boundary is relatively small (class probabilities are anchored around ffi0.5)
- Results of both model ensembles were worse than the baseline (GCN). However, low correlation between outputs of the ensemble models implies that late fusion should be further investigated
- High GCN + HAN 2 AUC score suggests that the classifier is relatively robust to the probability cut-off threshold
- The GCN baseline was the best performing model, with an accuracy of 87% and AUC of 0.88

Discussion

- The structural data of news article propagation on social networks allows for more accurate models than those trained on the textual content of the news articles
- While social network data is a sensitive data source in the context of Deep Learning, for this poster used only fully anonimized data, and the trained model classifies article nodes rather than user nodes as either fake or real
- In future experiments, if larger datasets with social network information are available, late-fusion technique should be compared with a classifier that takes article embeddings (from HAN) and user embeddings (from GCN) as input

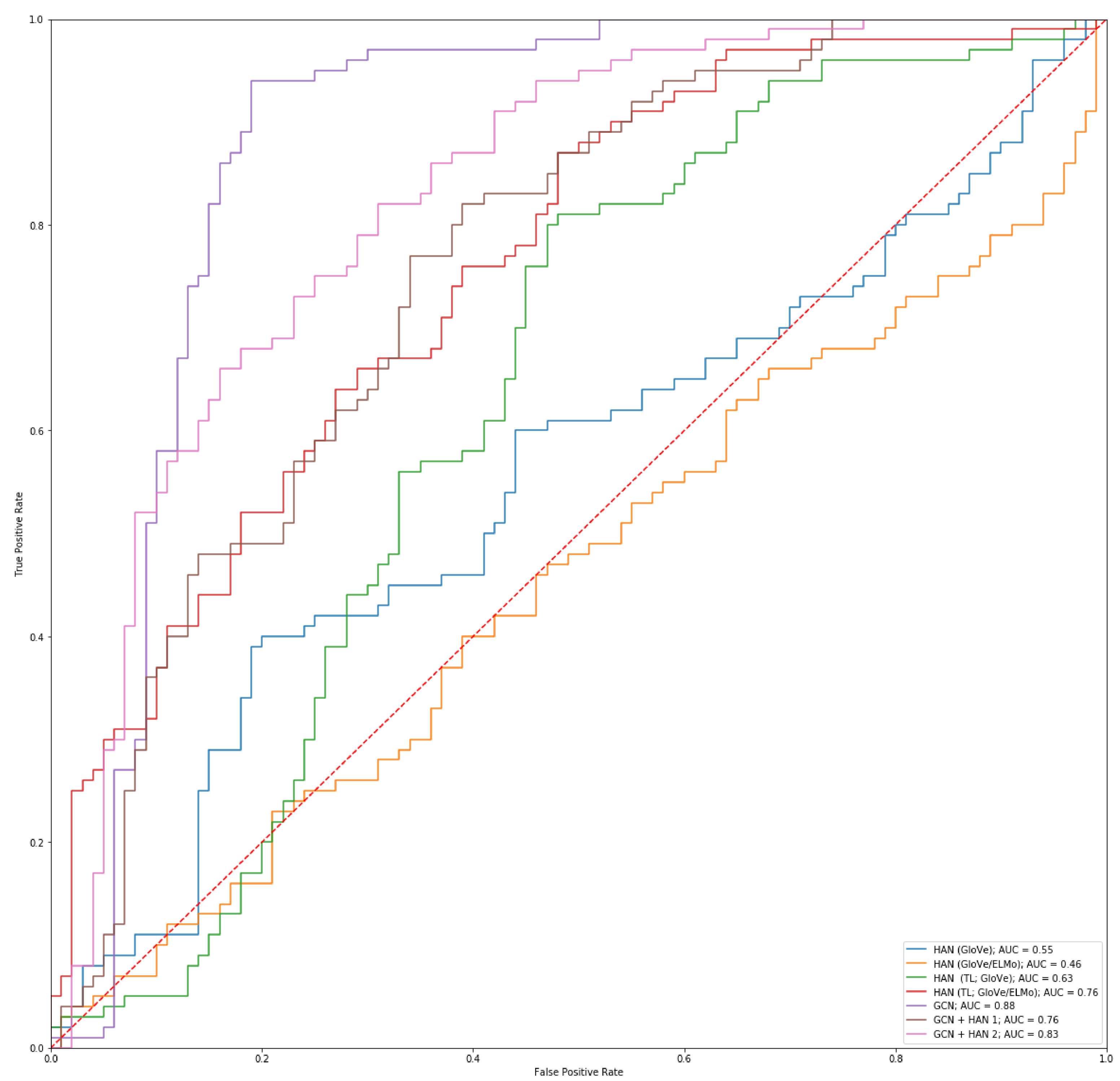

Figure 2. AUC curves for the (Multimodal) models trained and tested

References

Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks.

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2018). Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016). Hierarchical attention networks for document classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 1480-1489).

Zhou, X., Zafarani, R., Shu, K., & Liu, H. (2019, January). Fake News: Fundamental Theories, Detection Strategies and Challenges. In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining (pp. 836-837).