

ABSTRACT

The Digital age has witnessed an explosion of online content, making it increasingly challenging to differentiate between reliable information and misleading clickbait. Additionally, with the increase in the usage of large language models for content writing it is even more challenging for the general user to differentiate between clickbait and genuine content. Existing clickbait detection models often work on rule-based techniques which lack the nuanced understanding of human semantic knowledge, making them vulnerable to sophisticated clickbait techniques. Our goal is to develop a novel neural network that incorporates this semantic knowledge to improve clickbait detection accuracy.

Keywords – Transformers, Cross Attention, Clickbait detection, Text Classification, Pytorch

METHODS

Utilizing ConceptNet, we extract semantic relationships between tokens in headlines and articles. Our architecture incorporates multiple blocks, each featuring an encoder-encoder structure. One encoder processes the headline, while another handles the article. By employing cross-attention, the headline encoder considers both its context and the encoded article representation, and vice versa. The resulting outputs from all blocks are concatenated and directed through Normalization Layers and Feed Forward Layers. Additionally, Residual Connections are introduced to facilitate gradient flow. The outputs of all the blocks are concatenated and then fed into a classification layer, determining the probability of the headline-article pair being categorized as clickbait.

In today's online world, telling apart real info from misleading clickbait is tough. Our research focuses on *attending* to the interdependencies, semantic nuances in the content and spots clickbait, helping everyone stay informed and safe online!

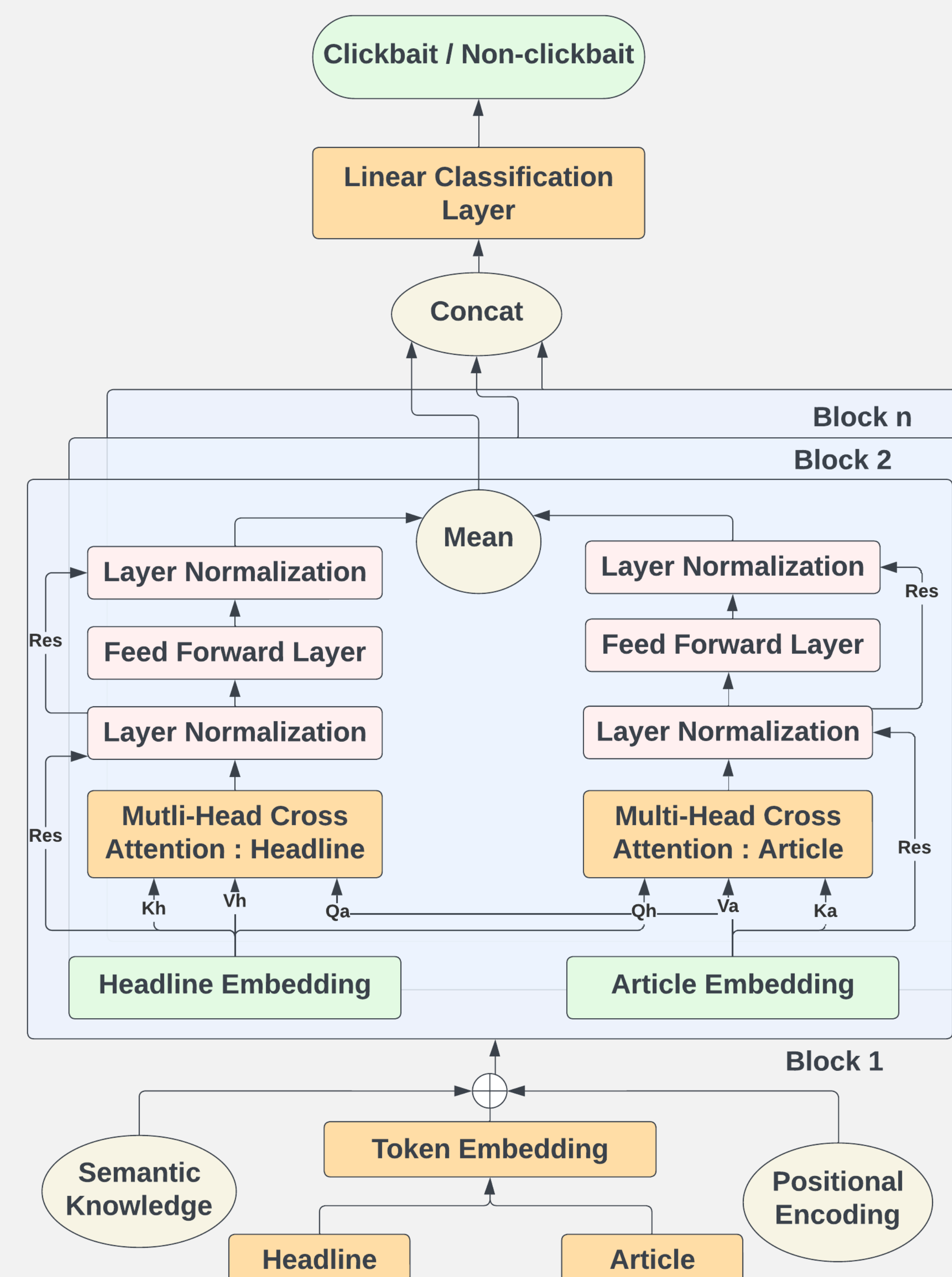


Fig. 1. High Level block Diagram

RESULTS

Results from using 3 different models are given below. First and second models use self attention with only headlines and articles, respectively. The third model, uses cross attention between headlines and articles.

Model	Accuracy	Precision	Recall	F1-Score
Headline only Encoder	80.8 %	73.8 %	57.3 %	64.5 %
Article only Encoder	83.2 %	75.6 %	60.2 %	67 %
Headline + Article Encoder	85.6 %	78.4 %	65.3 %	71.6 %

Table 1: Results of Clickbait Challenge Dataset

CONTACT INFORMATION

Lokesh Meesala
College of Computing and Software Engineering
meesala.lokesh@gmail.com
www.linkedin.com/in/lokesh-meesala

REFERENCES

- F. Wei and U. T. Nguyen, "An Attention-Based Neural Network Using Human Semantic Knowledge and Its Application to Clickbait Detection," in IEEE Open Journal of the Computer Society, vol. 3, pp. 217- 232, 2022, doi: 10.1109/OJCS.2022.3213791.
- FNC challenge - 2017 (www.fakenewschallenge.org/)
- Clickbait challenge - 2017 (www.clickbaitchallenge.org/)
- Speer, R., Chin, J., & Havasi, C. (2017). ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. Proceedings of the AAAI Conference on Artificial Intelligence, <https://doi.org/10.1609/aaai.v31i1.11164>.