Boosting Clickbait Detection through Semantic Insights and Attention-Driven Neural Network

Lokesh Meesala
College of Computing and Software Engineering
Kennesaw State University, Kennesaw GA
lmeesala@students.kennesaw.edu

***Abstract*—The digital age has witnessed an explosion of online content, making it increasingly challenging for users to differentiate between reliable information and clickbait, which is often misleading or sensationalized. Clickbait contributes to the spread of misinformation, phishing attacks, and illegal marketing practices, and manipulates users' decisions. Even from a business standpoint a clickbait might not lead to a conversion, A user might land on the page by following a clickbait and get frustrated and close the page. 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. As a result, clickbait detection has become an important research topic. 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 an easily configurable transformer neural network that incorporates this semantic knowledge to improve clickbait detection accuracy.**

I. Introduction

Clickbait is a deceptive online tactic designed to lure users into clicking on hyperlinks that ultimately lead to content of little value or relevance. Clickbait takes various forms, often using sensational headlines, provocative images, or misleading descriptions to misuse users' curiosity. It also undermines the integrity of digital platforms and online interactions. The rapid increase of clickbait has led to a host of negative consequences that extend beyond individual users to impact society as a whole. Its prevalence not only contributes to the spread of misinformation and phishing attacks but also gives rise to nefarious activities.

Clickbait is frequently used in illegal marketing practices to promote products or services through deceptive means, violating ethical and legal marketing standards. It can lead users to websites that violate their privacy by collecting personal data without consent, leading to privacy breaches. It can undermine the credibility of genuine news sources and content creators, making it harder for users to trust any online information. Clickbait often directs users to fictitious news articles, a situation particularly alarming in the realms of politics, health, or public safety. Relying on sensational headlines or exaggerated thumbnails, clickbait creates false expectations of extraordinary or shocking content, which rarely aligns with reality. Consequently, clickbait content, be it articles or videos, frequently disseminates inaccurate or deceptive information, significantly adding to the broader issue of misinformation on the internet. The constant exposure to clickbait can overwhelm users with information, making it difficult for them to distinguish between valuable content and sensationalized or misleading material. It can impair decision-making abilities and hinder the ability to critically evaluate information.

In recent times, the rise of large language models has further complicated this landscape. These advanced AI models are now being used for Automated Content Creation, which streamlines content creation process and empower malicious actors to mass-produce clickbait, using sensational headlines for capturing users' attention. Produce vast content, increasing online clickbait. This flood normalizes deceptive headlines, amplifying misleading claims users encounter. Large language models can also be used to craft targeted clickbait by analyzing data, appealing to users' emotions and biases, enhancing engagement. They can be finetuned to automatically adjust the clickbait creation process based on audience responses, making detection challenging due to evolving techniques.

Existing clickbait detection models often rely on rule-based techniques, offering some detection capability but lacking the nuanced understanding of human semantic knowledge. This gap makes them vulnerable to sophisticated clickbait methods, highlighting the need for more advanced detection approaches.

Motivated by these challenges, our research aims to develop an innovative neural network approach. This method incorporates a deep understanding of human semantics to enhance clickbait detection accuracy significantly. By delving into the intricacies of human semantics, our model can effectively distinguish between clickbait and genuine content, providing a more reliable solution against deceptive online practices. Our contributions are as follows:

- Utilize ConceptNet [16] knowledge graphs, enabling the extraction of relations between words in headlines and articles.
- Apply human semantic knowledge to guide the attention mechanism, aiding in the formulation of keys, queries, and value vectors.
- Implement the Encoder-Encoder transformer architecture, employing cross-attention between headline and article encoders. This enables the model to learn complex semantic relationships between headlines and their corresponding articles.

II. Related Work

Researchers have explored diverse approaches to tackle the clickbait challenge. Some have delved into hidden features like user behavior patterns [1], while others have crafted domain-specific features [2]. Certain studies devised probabilistic models to address multi-truth discovery, acknowledging that a single content piece can contain multiple positive and negative claims [3]; they have reported an average F1-score of 0.87. Additionally, methods incorporating statistical and linguistic features, such as content format analysis, informal word usage detection, and simple similarity scores between headline and article words, have been applied [4] which gave the weighted average F1-score of 0.74. Another technique involves determining the stance of the headline concerning the article through lemmatisation-based n-gram matching [5] and achieve combined classifier accuracy of 0.89.

The above mentioned methods lack targeted methods to block carefully crafted clickbait articles, inability to effectively add non-textual features such as images and videos, and user comments on articles. In the stance detection research they acknowledge that it it not directly applicable to the real world scenario since the data is artificially boosted by randomly combining headlines and article bodies to create unrelated headline/article pairs datapoints. Some researchers have created browser based extensions collect information on users interests and offer personalized classification by training the models based on word length, length of the syntactic dependencies, usage of hyperbolic words and internet slang's [6]. All of the above mentioned works have used supervised machine learning models such as Naive Bayes, Support Vector Machines, XGBoost and Light Gradient Boosting Machines. Clickbait articles are known to originate and spread form various social media accounts. Extracting features to analyze the behavior of the user account through deep learning [11] is also used as a mechanism to trace and stop the spread of clickbait articles.

Since all the above methods require intense feature engineering and are also difficult to adapt for ever-changing landscape of clickbait creation. Deep learning techniques have became popular for text classification in general and also used for clickbait detection. Domain Rating Filters, Whitelist and Blacklist labels [7] are used in the RNN models to efficiently distinguishing between legitimate and illegitimate links. A novel blockchain technology layer is used to validate the authenticity of the sources [8]. Knowledge-graphs are created using graph convolutional network and used in the clickbait detection which also provides explainability by tagging terms and phrases that could indicate exaggeration, emotional imbalance, special sentence patterns [12]. Creating a useful knowledge graph for clickbait detection needs a good domain specific training data. The invention of transformer architecture and attention mechanisms changed the landscape of natural language processing. It is being used in many language tasks including clickbait detection. Researchers have adopted the transformer architectures and fine-tuned it for clickbait detection [9] [10].

Wei et al. proposed the use of human semantic knowledge to drive attention mechanism [13]. Wordnet is used to extract the relation between words in the headline and the article which is used to create the headline-article representations. These vectors are passed into the BiGRU layers for further classification which showcased significant improvements in accuracy, precision, recall, and F1 scores, achieving approximately 75%. The enhancements made by incorporating WordNet for semantic knowledge resulted in a 3% to 5% performance boost across pre-trained models like BERT, RoBERTa, and XLNet when tested on the clickbait challenge [14] and Fake News Challenge [15] datasets.

In our research we intend to use the transformer encoder-encoder architecture to learn the headline-article representations. Additionally, ConceptNet [16] provides more relations to be explored serving as a potential alternative to the use wordnet to extract human semantic knowledge. This knowledge is used to create the key, query and value vectors in the model architecture.

III. Methodology

Our research adopts an advanced encoder only architecture. In this setup, each block contains two types of layers to perform attention mechanism. In the self attention layer each element in the input sequence attends to every other element, assigning different attention weights based on their relevance. This self-awareness allows the model to weigh the importance of each token with respect to others, capturing intricate patterns and long-range dependencies within the sequence. On the other hand, a cross-attention layer extends this mechanism to handle interactions between headline and article sequences. It enables the model to selectively focus on relevant information from one sequence while processing another, facilitating the extraction of meaningful relationships between the tokens. By aligning the attention mechanism across multiple sequences, cross-attention contributes to the creation of comprehensive and contextually rich representations is used to create self aware vectors. Figure 1 represents the high level components involved while Figure 2 represents each attention layer. Each attention layer has multihead architecture, followed by a feed forward layer with layer normalization's between them. Residual connections allow for uninterrupted gradient flow.

$$SelfAttention(Q_h, K_h, V_h) = softmax(\frac{Q_h \cdot (K_h)^T}{\sqrt{d_k}}) \cdot V_h$$

$$MultiHead\_SA = SA_1 \oplus SA_2 \oplus \ldots \oplus SA_n$$

$$SA\_output = LN(FF(LN(MultiHead\_SA)))$$

$$CrossAttention(Q_a, K_h, V_h) = softmax(\frac{Q_a \cdot (K_h)^T}{\sqrt{d_k}}) \cdot V_h$$

$$MultiHead\_CA = CA_1 \oplus CA_2 \oplus \ldots \oplus CA_n$$

$$CA\_output = LN(FF(LN(MultiHead\_CA)))$$

To enhance the model's understanding of the textual content, we harness the semantic knowledge from ConceptNet [16]. We used the semantic vectors from numbersbatch [17]; which is a part of the ConceptNet open data project. It is built in a 300 dimension vector space using an ensemble that combines data from ConceptNet, word2vec, GloVe, and OpenSubtitles. For our project we needed vectors of 120 dimension, so we have applied PCA for dimentionality reduction. we create the headline_input and article_input by adding respective the learnable embedding vectors, conceptnet knowledge vectors and positional encoding vectors creating a list of two inputs.

This input list is passed into the blocks in which each input element is passed into the self attention layers separately. Inside the self attention layer, same copy of inputs are passed into each head of the multi-head self attention layer. In each head, key, query and value vectors are created and attention computations are performed. The output of all the heads are concatenated and sent to the feed-forward layer which adds non-linearity to the model. To make sure that all the layer outputs are in same scale, multiple layer normalization layers are introduced before and after the feed-forward layer. The same steps are performed for headline and article inputs separately.

The self attended inputs are sent to cross attention layer. The same steps from previous self attention layer are performed in this layer too, difference is that the keys, values and queries are created from the self attended inputs rather than raw inputs like before and the query from headline is send into cross attention heads of the article and vice verse.

For preparing the data we used bert-base-uncased [18] model from hugging face to convert the raw text into tokens. In the given dataset about 75% of text sequences have less than 640 tokens. So we choose this threshold as the max length for truncating the text. A custom collate function is used for padding the text based on the max length sequence of a given batch.

By utilizing the knowledge base, self attention, cross attention mechanisms, intricate interdependencies and semantic nuances present in both the headline and article are extracted. These outputs are averaged and concatenated. The size of this vector will twice the size of embedding dimension which is 240. This is sent to a classification head, which determines the probability of the headline-article pair being clickbait or not.

IV. Experimental setup

A. Dataset

We used the publicly annotated dataset for training and evaluating the performance of the proposed model.

Webis Clickbait Corpus: This is the benchmark dataset used in clickbait detection challenge [14]. It contains a total of 40,976 articles along with their headlines. All these articles are from twitter posts that had been published between November 2016 and June 2017. These articles are annotated using Amazon Mechanical Turk where each article is rated on a 4-point scale [not click baiting (0.0), slightly click baiting (0.33), considerably click baiting (0.66), heavily click baiting (1.0)].

Fig. 1. High Level block Diagram

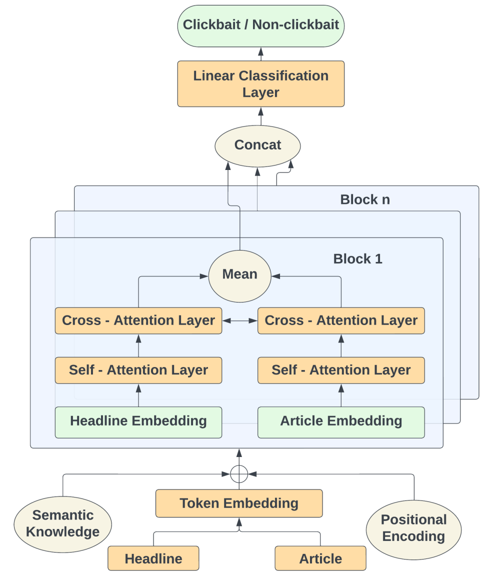

Fig. 2. Attention Layers

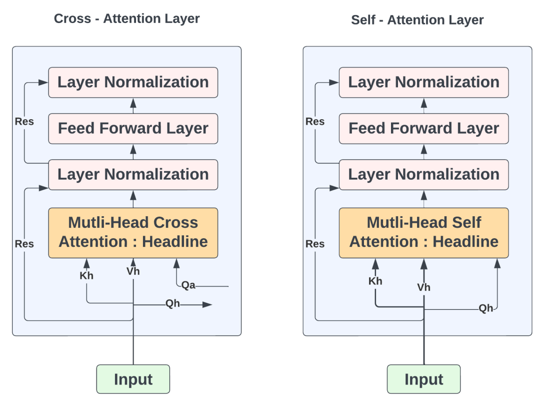

TABLE I
Data Distribution

Dataset	Clickbait	Non-Clickbait
Webis Clickbait Corpus	7953 (30%)	17245 (70%)

TABLE II
Data Statistics

Dataset	Avg. Headline Length	Avg. Article Length
Webis Clickbait Corpus	12	575

In our implementation We considered the threshold of 0.5 to categorize the data into clickbait and non clickbait articles. The corpus is divided into two parts, a training and a test dataset with a 70% split. The tables 1 and 2 summarize the distribution and statistics of the dataset.

TABLE III
MODEL PARAMETERS

Parameter	Value
Max Length	640
Number of Blocks	4
Embedding Dimension	120
Number of Heads	4
Dropout	0.2
Learning rate	0.0003
Batch Size	64
Epochs	30

TABLE IV
CLICK BAIT CORPUS RESULTS

Model	Accuracy	Precision	Recall	F1-Score
BertForSequenceClassification	0.812	0.73	0.70	0.714
Custom Model (Headline only)	0.774	0.615	0.683	0.647
Custom Model (Article only)	0.815	0.725	0.628	0.673
Custom Cross Attention Model	0.836	0.736	0.717	**0.727**

B. State of the Art Models

The BERT (Bidirectional Encoder Representations from Transformers) model [18] for sequence classification is a pre-trained natural language processing (NLP) model capable of understanding and classifying sequences of text. Leveraging bidirectional context analysis, BERT captures intricate language relationships by considering the complete context of each word in a sequence, resulting in highly effective and contextually rich representations for tasks like sentiment analysis, named entity recognition, and other forms of sequence classification. We used the BertForSequenceClassification model from hugging face transformers package for comparing the results with our model.

C. Evaluation Metrics

Given the nature of this problem as a binary classification task, various metrics including accuracy, F1-score, precision, and recall are employed. Achieving a delicate balance between precision and recall is essential in clickbait detection applications. For instance, it is vital to avoid mis-classifying genuine news articles as clickbait (high precision), while simultaneously capturing as many actual clickbait posts as possible to prevent the spread of false news and user dissatisfaction (high recall). To reconcile these objectives, the F1-score emerges as the most suitable metric for striking an appropriate balance between precision and recall in these contexts

D. Implementation Details

We choose to implement the encoder-encoder architecture using pytorch. After conducting various experiments we choose the parameters given in table 3 for our model. Our system configuration is as follows, 13th Gen Intel Core i9-13900HX processor with GeForce RTX 4080 GPU and 32 GB RAM.

E. Results

Our results are provided in Table 4. The Bert Model gave descent F1-Score of 71.4%. The headline and article are sent to the Bert Model, which uses only self attention to classify the headline-article pairs as clickbait or not. It also did not use the conceptnet embeddings. Also, given the limited training examples, self attention alone is not sufficient to extract all the inter dependencies in the text. The second and third row represent the results using the custom model with self attention using the headline and article individually. Our cross attention model outperformed both the models which shows that cross attention model was useful in extracting the relations. All the 3 models used the conceptnet embeddings.

F. Ablation Study

1) Impact of Embedding Size on Attention Mechanisms: Figure 3 shows the impact of using different embedding sizes. All the parameters were fixed and we ran the experiments for embedding sizes from 80 to 200. The embedding size of 120 gave the best results. The choice of embedding size depends on the size of the training data used. A smaller data set will not have enough context that can be used by the embedding dimension to learn the relationships and dependencies. On the other hand, If we use lower embedding size for a large data set then there wont be enough dimensions to represent the contextual details of the training data.

2) Impact of Head Sizes and Block Sizes: Number of heads and number of blocks (layers) play an important role in attention mechanism. Each head can focus on specific semantics in the given text. In our study, we have experimented with different head sizes from 3 to 6, and different blocks sizes from 2 to 4 without exceeding our system configuration. Each of them took different number of epochs to reach the optimal minima before early stopping. Head Size 4 and Block Size 4, gave the best F1 Score of 72.7%. The reason could be rooted in the intricate dynamics of attention mechanisms within the neural network architecture. In our study, we observed that a head size of 4 facilitated a diversified exploration of semantics in the given text. Each attention head specializes in capturing distinct aspects and patterns within the data, enabling the model to effectively attend to multiple dimensions simultaneously. Furthermore, a layer size of 4 provided the model with the necessary depth to hierarchically learn complex representations from the input data. The combination of both the head and layer sizes allows the model to strike a balance between capturing fine-grained details and learning high-level abstractions without over-fitting. Figure 5 shows the impact of using different sizes in the model performance.

3) Impact of Cross Attention: We have evaluated the model in 3 different input setting: Using only the headline, Using only the article; both of these use self-attention only. Finally we have used both the headlines and articles along with cross attention. Figure 4 illustrates the results, which indicates that the third input configuration yielded the most favorable outcomes. This underscores the significance of cross attention in enhancing model performance, as it evidently played a

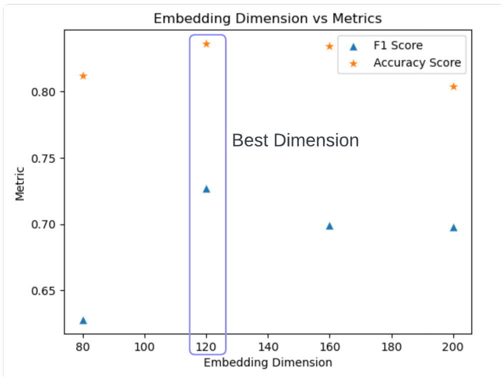

Fig. 3. Embedding Dimension Study

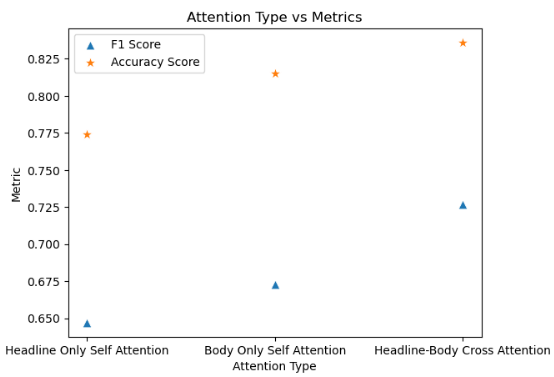

Fig. 4. Attention Mechanism Study

crucial role in achieving superior results compared to the other input configurations.

V. Conclusion

In conclusion, our comprehensive exploration into clickbait classification models has uncovered valuable insights regarding the role of attention mechanisms, embedding sizes, and model configurations. The initial utilization of the Bert Model demonstrated respectable performance, but its reliance solely on self-attention proved inadequate for capturing the nuanced dependencies within limited training examples. Our subsequent custom models, incorporating cross-attention followed by self-attention, showcased better results by extracting meaningful relations between headlines and articles. Notably, the inclusion of conceptnet embeddings consistently contributed to improved model performance across all configurations.

Further analyses shed light on the nuanced impact of embedding size, emphasizing the delicate balance required for effective representation based on the dataset's scale. The investigation into head sizes and block sizes within the attention mechanism provided valuable insights into the optimal configuration, with a head size of 4 and a block size of 4 yielding the best F1 Score.

Finally, our model has easy configurability, accommodating various sequential input variants such as text and video. This

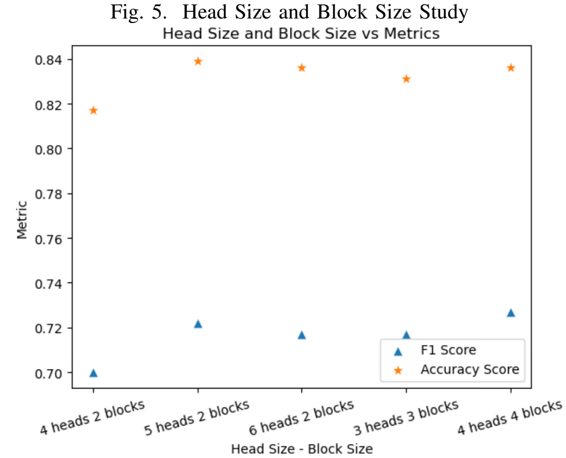

Fig. 5. Head Size and Block Size Study

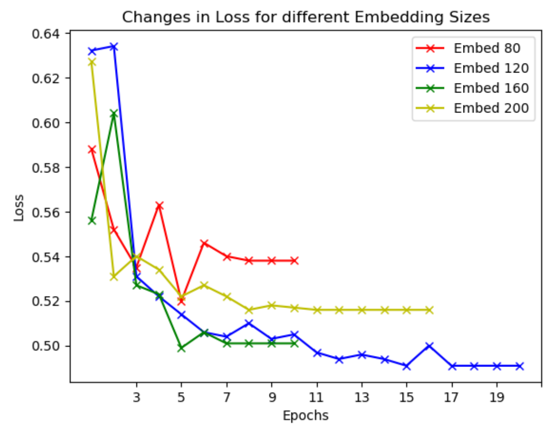

Fig. 6. Changes in loss function for different embedding sizes.

versatility positions our model as a valuable tool for clickbait detection, not only in traditional textual content but also in dynamic platforms like social media, including YouTube. This adaptability makes our model a promising asset for content moderation and user protection in the dynamic and multimedia-rich environment of social media platforms.

References

[1] Yimin Chen, Niall J. Conroy, and Victoria L. Rubin. 2015. Misleading Online Content: Recognizing Clickbait as "False News". In Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection (WMDD '15). Association for Computing Machinery, New York, NY, USA, 15–19. https://doi.org/10.1145/2823465.2823467

[2] Indurthi, V., Oota, S. R. (2017, October 8). Clickbait detection using word embeddings. arXiv.org. https://arxiv.org/abs/1710.02861.

[3] Xianzhi Wang, Quan Z. Sheng, Lina Yao, Xue Li, Xiu Susie Fang, Xiaofei Xu, and Boualem Benatallah. 2016. Truth Discovery via Exploiting Implications from Multi-Source Data. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management (CIKM '16). Association for Computing Machinery, New York, NY, USA, 861–870. https://doi.org/10.1145/2983323.2983791

[4] Prakhar Biyani, Kostas Tsioutsiouliklis, John Blackmer. 8 Amazing Secrets for Getting More Clicks: Detecting Clickbaits in News Streams Using Article Informality Proceedings of the AAAI Conference on Artificial Intelligence, 30 (2016) .

[5] Peter Bourgonje, Julian Moreno Schneider, and Georg Rehm. 2017. From Clickbait to Fake News Detection: An Approach based on Detecting the Stance of Headlines to Articles. In Proceedings of the 2017

EMNLP Workshop: Natural Language Processing meets Journalism, pages 84–89, Copenhagen, Denmark. Association for Computational Linguistics.

[6] Chakraborty, A., Paranjape, B., Kakarla, S., Ganguly, N. (2016, October 31). Stop clickbait: Detecting and preventing clickbaits in online news media. arXiv.org. https://arxiv.org/abs/1610.09786

[7] Razaque, A.; Alotaibi, B.;Alotaibi, M.; Hussain, S.; Alotaibi, A.;Jotsov, V. Clickbait Detection Using Deep Recurrent Neural Network. Appl. Sci. 2022, 12, 504. https://doi.org/10.3390/app12010504

[8] A. Razaque et al., "Blockchain-Enabled Deep Recurrent Neural Network Model for Clickbait Detection," in IEEE Access, vol. 10, pp. 3144-3163, 2022, doi: 10.1109/ACCESS.2021.3137078.

[9] M. Glenski, E. Ayton, D. Arendt, and S. Volkova, "Fishing for clickbaits in social images and texts with linguistically-infused neural network models," 2017, arXiv:1710.06390.

[10] Y. Zhou, "Clickbait detection in tweets using self-attentive network," 2017. arXiv:1710.05364.

[11] Sahoo, S.R.; Gupta, B.B. Multiple features based approach for automatic fake news detection on social networks using deep learning. Appl. Soft Comput. 2021, 100, 106983. [CrossRef] 20. Khan, M.A.; Kim, J. Toward developing efficient Conv-AE-based intrusion detection system using heterogeneous dataset. Electronics 2020,

[12] Zhou, M., Xu, W., Zhang, W. et al. Leverage knowledge graph and GCN for fine-grained-level clickbait detection. World Wide Web 25, 1243–1258 (2022). https://doi.org/10.1007/s11280-022-01032-3

[13] 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.

[14] "FNC challenge," 2017. [Online]. Available: http://www.fakenewschallenge.org/

[15] "Clickbait challenge," 2017. [Online]. Available: https://www.clickbait-challenge.org/

[16] 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, 31(1). https://doi.org/10.1609/aaai.v31i1.11164.

[17] Commonsense. (n.d.). Commonsense/conceptnet-numberbatch. GitHub. https://github.com/commonsense/conceptnet-numberbatch

[18] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. CoRR, abs/1810.04805. https://arxiv.org/abs/1810.04805