Knowledge-Aware Document Representation for News Recommendations considering image and sentiment analysis of news

Reza Barati
Politecnico di Torino
Turin, Italy
Reza.Barati@studenti.polito.it

Faezeh Saeedian
Politecnico di Torino
Turin, Italy
Faezeh.Saeedian@studenti.polito.it

Sarina Takalloo
Politecnico di Torino
Turin, Italy
Sarina.Takalloo@studenti.polito.it

***Abstract*—Personalized news recommendation has become increasingly important in the era of digital information. To generate effective recommendations, it is critical to represent news articles accurately and efficiently. In this work, we introduce Extended KRED, which builds on the Knowledge-Aware Document Representation for News Recommendations (KRED) by adding three extensions to the original approach. These extensions include comparing different sentence transformers, incorporating image data from the V-MIND dataset, and using semantic analysis to consider article sentiment. We demonstrate how these extensions affect the performance of category classification model of KRED on several evaluation metrics. The code is available at: https://github.com/faezesaeedian/KredVertClassify.git**

***Index Terms*—Sentiment analysis, News Classification, Deep NLP**

I. Introduction

The abundance of digital news sources has made it increasingly difficult for users to keep up with the news that matters to them. Personalized news recommendation systems have emerged as a promising solution to this problem by providing users with tailored news articles based on their interests and preferences [1]. However, to generate effective recommendations, it is critical to accurately and efficiently represent news articles. However, in the domain of news, there are certain peculiarities that need to be considered, such as the time-sensitive nature of news and the importance of identifying relevant entities. One promising research direction is the use of a knowledge graph to enhance news document vectors by fusing them with knowledge entities. Liu Danyang et al. [2] have proposed a Knowledge-Aware Representation Enhancement model for News Documents that utilizes an entity representation layer, a context embedding layer, and an information distillation layer to produce a representation vector that can be used in various downstream tasks. In this context, KRED was proposed as an approach to represent news articles using external knowledge sources. KRED leverages knowledge graphs to capture the semantic relationships between news articles and their corresponding entities, such as people, organizations, and locations. This enables KRED to better capture the meaning and context of news articles, leading to improved recommendation performance. Building on the success of KRED, we propose Extended KRED, which adds three extensions to the original approach.

- First, we compare different sentence-transformers to evaluate their effectiveness in representing news articles. Specifically, we compare BERT with RoBERTa and XLM model to select the best-performing model for document embedding.
- Second, we incorporate image data from V-MIND [3], which is an extended version of the MIND [4]dataset that includes news pictures. This extension enables us to match news articles with relevant images, expanding and diversifying the dataset.
- Third, we use a sentiment analysis to consider the effect of positive or negative context of news in classifying them.

In this report, we describe the Extended KRED approach and evaluate its performance compared to the original KRED.

II. Methodology

KRED is a knowledge-aware representation enhancement model that takes into account knowledge entities when processing text, unlike other models like BERT. The incorporation of knowledge entities is crucial for recommendation tasks, particularly in the context of news that contain references to people, places, events, and other entities that provide important context and information. KRED achieves this by creating a knowledge-enhanced document vector that can be used in various applications. The model comprises three essential components: the entity representation layer, the context embedding layer, and the information distillation layer.

The entity representation layer embeds information about the relations with other entities, using TransE [5]to learn representations for each entity and relation. The context embedding layer extracts entity information related to the document and embeds it into its representation by using position, frequency, and category encoding. The information distillation layer uses an attentive mechanism to highlight the most important entities within an article, summarizing the attention-weighted entity representations and concatenating the result with the original document vector (VD) to obtain the final Knowledge-aware

Document Vector (KDV). In our study, we extend the KRED model by adding features of image and sentiment in the representation vector of each news and compare different sentence embedding models to find the best one and evaluate its performance under these conditions. Due to the computational cost and resource limitations, we selected the article category classification task and focused to expand it. Note that the default task of the KRED baseline is user-to-item classification. The hyperparameters used for the selected task are listed in Table I.

A. Proposed Extensions

Before discussing the extensions, it is important to understand the structure of the dataset used in this study. The dataset used in this study is organized in the form of a click log, where each user click within the portal is logged and saved. A unique hash code is assigned to each clicked item, allowing for item-level identification. The hash code for each clicked article is then extracted and mapped to the corresponding title, category, subcategory, link, and description. In our study, we utilized the Microsoft Inner Loop News Dataset (MIND) as the basis for our experiments. The dataset consists of rows containing an identification number, category, subcategory, title, the website link and entities for each news article. We employed MIND as the foundation for both our baseline model and all subsequent extensions evaluated in our study.

1) Transformer Comparison: The KRED model utilizes BERT as a transformer for generating a Knowledge-enhanced Document Vector (KDV) from an arbitrary Document Vector (DV), denoted as v_d. The transformer's ability to capture long-term dependencies and contextual information in text data is a significant advantage in various NLP tasks. The KDV produced can then be used in downstream applications. Unlike traditional NLP models, transformers do not rely on hand-crafted features or domain-specific knowledge. Instead, they learn representations of text data in an unsupervised manner, by processing large amounts of text data. This pretraining step allows the model to capture general knowledge about language and context, which can then be fine-tuned on specific tasks with smaller amounts of labeled data. In this extension, we tried to compare BERT with two other transformers, RoBERTa and XLM, which are advanced versions of BERT with some improvements. The advantages of these models compared to BERT are:

- More training data: RoBERTa is trained on an even larger corpus than BERT, while XLM transformer is trained on a diverse set of texts with multiple formats, including HTML and XLM.
- Pre-training approach: RoBERTa uses a more effective pre-training approach than BERT, which includes removing the next sentence prediction task and training on longer sequences.
- Dynamic Masking: RoBERTa and XLM transformer use dynamic masking, which means that they mask out different tokens during training, as opposed to the static masking in BERT.
- Better performance: RoBERTa and XLM transformer have achieved state-of-the-art performance on many NLP tasks, indicating that they are more effective than BERT in capturing language semantics and structures.

in this experiments, we implement Roberta and XLM transformers and evaluate the performance compare to BERT. The results of our evaluation are presented in Table II.

2) Enhancing News Embedding with Image Data: In this study, we investigated the impact of incorporating images into news articles by using the V-MIND dataset, which is an extension of the MIND dataset with associated news images. Each news article in V-MIND is linked to a corresponding image identified by a shared news ID. To incorporate the images into our model, we first extracted image features using the VGG16 network, which outputs a vector with a dimension of 100, that describe the image content. These features were saved in a dictionary for each news ID. The reason for selecting VGG16 [6] network is to mitigate the computational costs that would be incurred with other more complex networks. for each news, we concatenated the extracted image features with document vector(DV), which is BERT transformer for baseline, and the output of third layer(Information Distillation Layer) to produce a new representation vector for each news and then, fed this vector to the category classification model to be trained. Fig. 1 illustrates the step-by-step process of the extension, highlighting each stage of data processing and model training. The results of this extension are shown in Table II.

3) Sentiment Analysis of the news: As previously mentioned, KRED produces a vector representation of each news using the entities extracted from the article's abstract and title. However, incorporating the sentiment of the news may aid the model's performance. To this end, we utilized sentiment analysis on the context of the news. The predicted sentiment outcome was then concatenated to document vector(DV), which is BERT transformer for baseline, and the output of third layer, which was subsequently used for training the KRED model. We utilized a pre-trained model for detect sentiment using Hugging Face Transformers Trainer. The process of the extension is elaborated in Fig. 1 which outlines the key steps involved in the implementation. The results of this extension are shown in Table II.

III. Experiment

A. Dataset

1) MIND: MIND dataset is a large-scale dataset for news recommendation research, containing behaviors and news articles from the Microsoft News website. The dataset is widely used for research purposes due to its large size and the diversity of the news articles. The dataset includes information about users, items, and entities, and is already split into train, test, and validation sets.

2) V-MIND: V-MIND is an extension of the MIND dataset. The dataset contains images and their associated news ID,

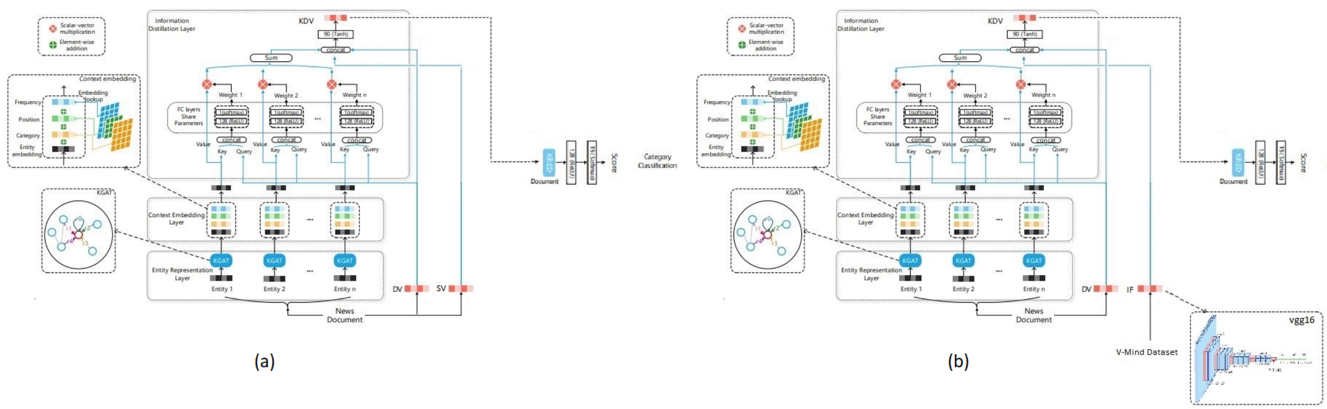

Fig. 1. (a) Framework of adding Sentiment Analysis of the news (b) Framework of adding Image Feature of News

making it suitable for multimodal research in news recommendation systems. V-MIND is based on the same user behavior data from the Microsoft News website as the MIND dataset.

B. Experimental Design

The experiments were conducted on a virtual machine of Google Colab Pro with an average runtime of 7.2, 7.2, 7.2, and 8.4 minutes per epoch for Baseline, Transformer Comparison, News Embedding with Image Data, and Sentiment Analysis, respectively. The average hardware specifications used were 12.1-25.5 GB RAM, 1.3-15 GB GPU, and 34.7-166.6 GB of disk space. The performance of the extended KRED in Category Classification task, was evaluated with Accuracy and F1-score Macro.

C. Results

The hyper-parameters used in the experiments for MIND and its implementations are presented in Table I. For the first extension, Transformer Comparison, RoBERTa outperformed BERT while XLM was found to be similar to BERT. This may be attributed to the advantages of RoBERTa discussed in section II-A.

After observing the behavior of second extension, we have come to the realization that the outcomes differed from our expectations. This may be attributed to inconsistencies between the pictures and news, as some pictures in V-MIND may not accurately depict the news they are associated with. For instance, a news article regarding weather category, be linked to a picture of a salad bowl, which may cause the model to misclassify it as a sports news article. Fig.2, shows more examples of mismatches.

Upon analyzing the results of third experiments, we concluded that the observed outcomes of the extension are due to the inherent difficulty in determining the polarity of news articles, particularly for articles that belong to categories that contain both positive and negative aspects, such as sports or weather. The model's performance was not improved by the incorporation of sentiment analysis effectively.

TABLE I
HYPER-PARAMETERS SET

Hyper-Parameters	values
Epochs	50
Batch size	64
Optimizer	Adam
Learning rate	0.00002
Weight decay	0.000001

TABLE II
RESULTS

Type of model	Accuracy	F1-score
Baseline	0.6709	0.2615
Transformer comparison - XLM	0.6661	0.2724
Transformer comparison - Roberta	0.6746	0.2717
Enhancing news embedding with image data	0.6254	0.1781
Sentiment analysis of news	0.6485	0.2262

IV. CONCLUSIONS

In this work, we proposed Extended KRED, an extension to the KRED approach for personalized news recommendation. The three extensions we added to the original approach were intended to improve the accuracy and efficiency of news article representation, but our evaluation results showed that these extensions did not always lead to better performance. Specifically, incorporating image data and using semantic analysis for article sentiment evaluation led to worse performance in terms of F1 score and accuracy compared to the original KRED approach. However, when we used Roberta sentence transformers instead of Bert, the performance of our proposed approach was improved, achieving better results in terms of F1 score and accuracy score.

Moreover, we tested the incorporation of extra information sources, including sentiment analysis and image data, to improve the representation of news articles. Although this approach appears to be promising, our findings suggest that its practical impact on the performance of the system may not be as significant as initially thought. One of the main reasons for

Fig. 2. Some example of mismatches between image and news category in four different categories

the lack of impact was due to the quality of the dataset we used for the images. We found that there were many inconsistencies between the real category of the news article and the image that was linked to it. Also, incorporating sentiment analysis did not effectively enhance the model's performance for articles that belong to categories that have both positive and negative elements.

References

[1] J. Lian, F. Zhang, X. Xie, and G. Sun, "Towards better representation learning for personalized news recommendation: a multi-channel deep fusion approach," in Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18, pp. 3805–3811, International Joint Conferences on Artificial Intelligence Organization, 7 2018.

[2] Danyang Liu, Jianxun Lian, Shiyin Wang, Ying Qiao, Jiun-Hung Chen, Guangzhong Sun, and Xing Xie. 2020. KRED: Knowledge-Aware Document Representation for News Recommendations. In Fourteenth ACM Conference on Recommender Systems (RecSys '20), September 22–26, 2020, Virtual Event, Brazil. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3383313. 3412237

[3] Songhao Han, Wei Huang, and Xiaotian Luan, VLSNR:Vision-Linguistics Coordination Time Sequence-aware News Recommendation". Submitted on 6 Oct 2022.

[4] F. Wu, Y. Qiao, J.-H. Chen, C. Wu, T. Qi, J. Lian, D. Liu, X. Xie, J. Gao, W. Wu, and M. Zhou, "MIND: A large-scale dataset for news recommendation," in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, (Online), pp. 3597–3606, Association for Computational Linguistics, July 2020.

[5] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," Advances in neural information processing systems, vol. 26, 2013.

[6] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv:1409.1556, 2014.