e-ISSN 2963-2870
Vol 5 2024

DEVELOPING RECOMMENDER SYSTEM FOR LEARNING MANAGEMENT SYSTEM USING TRANSFORMER AS ATTENTION-BASED MODELS

IKHSAN[1], KANI[2], MUHAMMAD ASYAR[3], MUSTARI[4], AKHMAD KHANAFI MASHUDI[5]

[1, 2, 4] *Universitas Terbuka Affiliation (INDONESIA)*
[3, 5] *Universitas Negeri Malamg (INDONESIA)*

ikhsan@ecampus.ut.ac.id

Abstract

In the era of information overload, recommender systems (RS) have become crucial tools in improving user experience across various domains, including Learning Management Systems (LMS). RS in LMS is particularly valuable for providing personalized learning recommendations that align with user individual preferences, abilities, and needs. This paper proposes to develop an RS for LMS using transformer models as attention-based systems. By leveraging the self-attention mechanism of transformers, the proposed system can accurately focus on the most relevant aspects of user interactions, resulting in more precise and relevant recommendations. Our experiments compare the transformer-based model with the Neural Collaborative Filtering (NCF) model, demonstrating the superiority of the transformer model in both HR@10 and NDCG@10 metrics. The transformer model achieves HR@10 of 70.59% and NDCG@10 of 50.34%, outperforming the NCF model by capturing more complex interactions between users and learning materials. The results highlight the potential of transformers to enhance personalized learning experiences in LMS, offering a more robust framework for understanding user behavior and delivering tailored learning content.

Keywords— recommender system, transformer, attention-based model

1 INTRODUCTION

In the era of information overload, recommender systems (RS) have become an essential tool in various domains, including Learning Management Systems (LMS). RS are widely used to help users find their desired items or services from a large collection of options. They can improve user satisfaction, loyalty, and retention, as well as generate revenue for the providers. However, designing effective RS is challenging, as they need to deal with complex and dynamic user preferences, item features, and system environments.

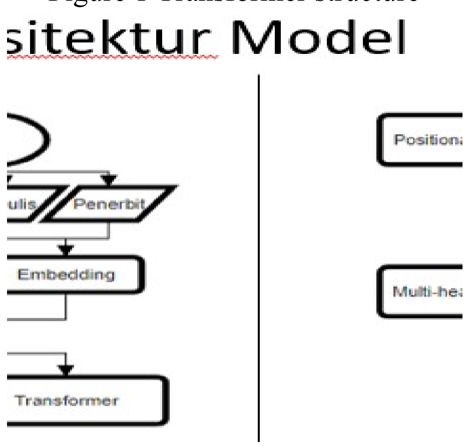
RS plays a crucial role in enhancing the user experience in the Learning Management System (LMS). According to [1], although LMS has facilitated access to various learning materials, the primary challenge is how to create a personalized and efficient learning experience for each user. They also state that many current LMS still function primarily as information repositories.

RS in LMS can provide recommendations for learning materials that align with individual interests, abilities, and needs by analyzing user learning patterns.

Technological advancements have enabled the application of attention-based transformer models as an RS. Transformers have become a trend in various fields of information technology due to their success in improving the performance of various applications, especially in generative language models like GPT (Generative Pre-trained Transformer). The success of these models in various fields has demonstrated their superiority in understanding context and generating more precise and personalized recommendations [2].

Transformers can analyze and understand user preferences more accurately with their excellent attention mechanism. This enables the system to provide more relevant recommendations. We hope that this work will not only contribute to the ongoing research in the field of transformer model and its application in RS but also inspire further innovation in the application of transformer in other fields.

Figure 1 Transformer structure



Related research about Transformers have become a focal point in various research areas related to recommender systems. Known as attention-based models, transformers have demonstrated their superiority in various applications, particularly in natural language processing and recommender systems. Research [3] introduced the transformer architecture, which underlies significant advancements in natural language processing and recommender systems. This study shows that the attention mechanism enables models to capture relationships between data more effectively and efficiently, forming the foundation for many modern transformer models used today.

One study demonstrates how BERT [4], a popular transformer model, can be used to improve the accuracy of recommender systems by understanding the context and sequence of user preferences. This model can capture complex relationships between recommended items and user preferences, resulting in more relevant and accurate recommendations.

Another study introduces the SASRec model [5], which uses the self-attention mechanism to more effectively capture the dynamics of user preferences in the context of recommender systems. The success of this model in enhancing recommendation performance underscores the significant potential of using transformers in developing better recommender systems.

Both studies [4], [5] indicate that transformer models, with their strong attention capabilities, can significantly improve the quality and accuracy of recommender systems. Therefore, the application of transformers in recommender systems is expected to positively impact the enhancement LMS performance.

The **model architecture** used in this research consists of embedding, transformer encoder, and a fully connected (FC) layer at the end of the model. The main function of the transformer encoder is to process the input sequence to generate contextual representations of the data. The use of the encoder is highly beneficial for understanding relationships between features in the data sequence in detail, allowing the model to capture complex latent relationships between the features. The main function of the transformer decoder is to generate output sequences and is typically used for generating multiple output, so it is not used in this model and is replaced by an FC layer to produce the final output. The use of the FC layer allows the model to remain simple yet effective in capturing complex patterns from the sequence data, enabling it to generate relevant recommendation probabilities.

Embedding is a technique in machine learning and deep learning used to convert categorical data into lower-dimensional numeric vectors. Embeddings help avoid the pitfalls of numerical ordering of data and capture complex latent relationships between users and items [6]. Embeddings enable recommender systems to make more personalized and relevant predictions. The learned embedding vectors can capture each user's unique preferences, allowing the model to provide more suitable recommendations.

Transformer is a neural network architecture designed to handle sequential data efficiently and effectively. Transformers use an attention mechanism to process sequential data, allowing the model to simultaneously focus on different parts of the input sequence and providing better ability to capture contextual relationships between elements within that sequence [3]. Fig. 1 shows the main components in the transformer model, which include: positional encoding, multi-head attention, and feed-forward networks.

e-ISSN 2963-2870
Vol 5 2024

Positional encoding is a crucial component of the transformer model that allows the model to understand the order of a sequence of information. Transformers process input in parallel, which requires a positional encoding method to recognize the position of tokens in the sequence so that the order information is preserved [3].

This transformer model uses sinusoidal positional encoding [3] to provide positional information to each token in the input sequence of the transformer model. The sinusoidal function is designed in such a way that each dimension of a token's position is filled with values from sinusoidal and cosine functions that have different periods.

For each position p, with i and d representing the dimension index and model dimension, respectively, for even positional encoding $PE_{(p,2i)}$ and odd positional encoding $PE_{(p,2i+1)}$, the following applies:

$$PE_{(p,2i)} = \sin\left(\frac{p}{10000^{\frac{2i}{d}}}\right) \tag{1}$$

$$PE_{(p,2i+1)} = \cos\left(\frac{p}{10000^{\frac{2i}{d}}}\right) \tag{2}$$

This approach has the advantage that the relative differences between the positions of two tokens can be explicitly represented through the encoding values, which are highly useful for understanding the order of information. The use of sine and cosine functions allows the model to easily learn the relationships between positions within the sequence using linear operations like dot products [3]. Sinusoidal positional encoding ensures that the transformer model can effectively process sequential information even while working in parallel.

Multi-head attention is a key component of the transformer model architecture, allowing the model to focus on different pieces of input information in a more flexible and efficient manner. Multi-head attention enables the model to capture various aspects of the relationships between words in a sentence by using several attention heads [3] in parallel. Each attention head works independently to extract different information from the input, and the results are combined to provide richer and deeper representations.

Multi-head attention processes the input using a scaled dot-product attention mechanism [3]. For Query (Q), Key (K), and Value (V) vectors, which represent the data being processed, and d_k, which is the dimension of K, the following applies:

$$\text{Attention Weights} = \text{softmax}\left(\frac{Q.K^T}{\sqrt{d_k}}\right) \qquad (3)$$

$$\text{Output} = \text{Attention Weights} \; . V \qquad (4)$$

In Multi-Head Attention, this process is not performed only once but in parallel with several heads, each having a different set of Query, Key, and Value vectors. Each head produces a different output, and these outputs are then combined and projected back to the desired dimensions.

The Feed-Forward Networks layer in the transformer serves to process and transform the input representation after the attention mechanism. Each layer in the transformer has a feed-forward network (FFN) that is applied separately at each token position in the input [3].

In FFN, there is a Multi-layer Perceptron (MLP) consisting of two linear layers with a ReLU activation function between them. The output of the MLP is then added and normalized using layer normalization [7]. Feed-forward networks allow the model to capture complex relationships in the data and process information efficiently.

The Fully Connected (FC) layer in the model architecture plays an important role in completing the prediction process after the data is processed by the embedding and transformer encoder sections. The FC layer at the end of the model serves to convert the feature representations generated by the transformer encoder into prediction scores that indicate how relevant a target is to a user's preferences. This process is carried out by fully connecting every neuron in the previous layer to the neurons in the next layer, allowing the model to leverage all the learned information.

In the transformer architecture, the decoder is typically used to generate output sequences such as text. However, in the case of this book recommendation model, there is no need to generate an output sequence. Therefore, the decoder function is replaced by an FC layer, which is more appropriate for the task of generating specific recommendation scores. By removing the decoder and using an FC layer, the model can stay focused on its primary task of providing relevant recommendations without dealing with unnecessary additional complexity.

e-ISSN 2963-2870
Vol 5 2024

2 METHODOLOGY

In the experiment setup, the data used for the study comes from XuetangX[2], one of the largest Massive Open Online Course (MOOC) platforms in China. XuetangX[2] provides a rich dataset that includes user interaction data with various courses, making it an ideal resource for tasks like recommendations and learning behavior analysis. This dataset was also employed in previous works, such as the Hierarchical Reinforcement Learning [8] which focuses on optimizing course recommendations by modeling user behavior at multiple levels.

During training, we create 4 negative interactions for each positive interaction by replacing the actual target with 4 random targets from the sample. In the testing phase, we pair each positive interaction with 99 random negative interactions from the sample [9]. We compare the transformer model with the neural collaborative filtering model [6] for each user and target pair to determine the probability of recommending a book to the user.

We use Hit Rate (HR) and Normalized Discounted Cumulative Gain (NDCG) as evaluation metrics in our experiments. Hit Rate (HR) measures how often relevant items appear in the top-N recommendation list. We use this metric for its simplicity and effectiveness in providing a quick overview of recommendation quality [6]. Normalized Discounted Cumulative Gain (NDCG) evaluates the quality of the recommendation ranking by considering the relevance of items and their positions in the list. We use NDCG for its ability to capture the ranking quality more comprehensively compared to HR [10].

Figure 2. Experiment results

Model Performances (%)

Metode	***HR@10***	***NDCG@10***
NCF	61.55	37.18
Transformer	70.59	50.34

We implement the model using PyTorch and run the code on the Kaggle platform with an NVIDIA P100 GPU, which offers high performance and efficiency in training deep learning models. We use the AdamW optimizer with a learning rate of 0.0001 and a weight decay of 0.001 to control overfitting. The loss function used is Binary Cross Entropy with Logits, suitable for binary classification tasks with logit output.

e-ISSN 2963-2870
Vol 5 2024

3 FUNDINGS AND DISCUSSION

Fig. 2 illustrates the overall performance of the two models compared: the proposed transformer model and the Neural Collaborative Filtering (NCF) model [6]. Based on the results, the transformer model outperforms the NCF model in the metrics HR@10 and NDCG@10.

The transformer model achieves an HR@10 value of 70.59% compared to 61.55%. This improvement is due to the transformer's ability to capture complex and long-term relationships in sequential data. The self-attention mechanism in the transformer allows the model to focus on the most relevant information at each input position, enhancing recommendation accuracy. The transformer model shows a significant improvement with a score of 50.34% compared to 37.18% for the NCF on the NDCG@10 metric. This is because the transformer's encoder effectively processes the sequential data of users and targets, producing rich and meaningful representations. This allows the model to better understand user preferences and target characteristics compared to the NCF approach, which uses a Multi-Layer Perceptron (MLP). The NCF model, relying on MLP to combine user and target embeddings, is less capable of capturing the more complex interactions. As a result, the recommendation performance of the NCF model is lower compared to the transformer model.

The transformer encoder captures complex relationships in user and book data, producing representations rich in information about user preferences. This process allows the model to understand the context and interactions between various features in data. The encoders enables the model to better capture the interactions between user preferences and target characteristics to capture more complex interactions, thereby enhancing the accuracy and relevance of the final recommendations.

4 CONCLUTION

Overall, the proposed transformer model demonstrates superior performance compared to the NCF model across all evaluation metrics. The significant improvement in HR@10 and NDCG@10 indicates that the transformer model is more effective in recommending relevant courses to users. This advantage is primarily due to the transformer's ability to capture complex and long-term relationships in sequential data, as well as the effectiveness of the attention mechanism employed. The reference study [6] provides important context for understanding this comparison.

ACKNOWLEDGEMENTS

For the first and foremost, Researcher would like to express their deepest gratitude for Almighty Allah SWT for his marvelous and amazing grace, for the countless blessings and love so the researchers have finally completed this subscribe as article. This full-text article is aimed to fulfill one of requirements for participating in INNODEL 2024. The research teams are also expressing extremely grateful to the Disruptive Learning Innovation flagship and Research Center of Excellence (PUI) Universitas Negeri Malang, ICE Institute dan LPPM Administrative Office Universitas Terbuka for supporting and opportunity.

REFERENCES

D. F. Murad, Y. Heryadi, B. D. Wijanarko, S. M. Isa, and W. Budiharto, "Recommendation System for Smart LMS Using Machine Learning: A Literature Review," in 2018 International Conference on Computing, Engineering, and Design (ICCED), 2018, pp. 113–118. doi: 10.1109/ICCED.2018.00031. [1]

J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," CoRR, vol. abs/1810.04805, 2018, [Online]. Available: http://arxiv.org/abs/1810.04805. [2]

A. Vaswani et al., "Attention Is All You Need," CoRR, vol. abs/1706.03762, 2017, [Online]. Available: http://arxiv.org/abs/1706.03762 [3]

F. Sun et al., "BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer," CoRR, vol. abs/1904.06690, 2019, [Online]. Available: http://arxiv.org/abs/1904.06690 [4]

W.-C. Kang and J. McAuley, "Self-Attentive Sequential Recommendation," in 2018 IEEE International Conference on Data Mining (ICDM), 2018, pp. 197–206. doi: 10.1109/ICDM.2018.00035. [5]

X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, "Neural Collaborative Filtering," CoRR, vol. abs/1708.05031, 2017, [Online]. Available: http://arxiv.org/abs/1708.05031 [6]

J. Alammar, "The illustrated transformer," The Illustrated Transformer--Jay Alammar--Visualizing Machine Learning One Concept at a Time, vol. 27, pp. 1–2, 2018. [7]

J. Zhang, B. Hao, B. Chen, C. Li, H. Chen, and J. Sun, “Hierarchical Reinforcement Learning for Course Recommendation in MOOCs,” Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, no. 01, pp. 435–442, Jul. 2019, doi: 10.1609/aaai.v33i01.3301435. [\8]

X. He, Z. He, J. Song, Z. Liu, Y.-G. Jiang, and T.-S. Chua, “NAIS: Neural Attentive Item Similarity Model for Recommendation,” IEEE Trans Knowl Data Eng, vol. 30, no. 12, pp. 2354–2366, 2018, doi: 10.1109/TKDE.2018.2831682. [9]

K. Järvelin and J. Kekäläinen, “Cumulated gain-based evaluation of IR techniques,” ACM Trans. Inf. Syst., vol. 20, no. 4, pp. 422–446, Oct. 2002, doi: 10.1145/582415.582418. [10]

e-ISSN 2963-2870
Vol 5 2024