

Article

A Multi-Level Location-Aware Approach for Session-Based News Recommendation

Xu Yu ^{1,2,3}, Shuang Cui ⁴, Xiaohan Wang ^{1,2,3}, Jiale Zhang ^{1,2,3}, Zihan Cheng ^{1,2,3}, Xiaofei Mu ^{1,2,3} and Bin Tang ^{5,6,*}

¹ Qingdao Institute of Software, China University of Petroleum (East China), Qingdao 266580, China; yuxu0532@upc.edu.cn (X.Y.); s24070049@s.upc.edu.cn (X.W.); z24070053@s.upc.edu.cn (J.Z.); s24070075@s.upc.edu.cn (Z.C.); z24070073@s.upc.edu.cn (X.M.)

² College of Computer Science and Technology, China University of Petroleum (East China), Qingdao 266580, China

³ Shandong Key Laboratory of Intelligent Oil & Gas Industrial Software, China University of Petroleum (East China), Qingdao 266580, China

⁴ School of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China; cuishuang@mails.qust.edu.cn

⁵ Qingdao Innovation and Development Base, Harbin Engineering University, Qingdao 266000, China

⁶ College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China

* Correspondence: tangbin@hrbeu.edu.cn

Abstract: Recently, personalized news recommendation systems have been widely used, which can achieve personalized news recommendations based on people's different preferences, optimize the reading experience, and alleviate the problem of information overload. Among them, session-based news recommendation has gradually become a research hotspot as it can recommend news without requiring users to log in or when their reading history is difficult to obtain. The key to session-based news recommendation is to use short-term interaction data to learn user preferences. Existing models often focus on mining news content information in sessions and do not fully utilize geolocation information related to news and sessions, and there is also a certain inconsistency between their training objective and model evaluation metric, leading to suboptimal model recommendation performance. In order to fully utilize geolocation information, this paper proposes a multi-level location-aware approach for session-based news recommendation (MLA4SNR). Firstly, a news-location heterogeneous graph is constructed, and a graph element-wise attention network is proposed to mine high-order relationships between news and location. Secondly, a session feature extraction network based on Transformer is proposed to extract session features. Then, a session-location heterogeneous graph is constructed, and a graph element-wise attention network is used to mine high-order relationships between sessions and locations. Finally, a loss function based on the NDCG is used to train the model. Experimental results on a real news dataset show that MLA4SNR outperforms the baselines significantly.

Keywords: session-based recommendation; news recommendation; location-aware; graph neural network; transformer

Academic Editor: Stefanos Kollias

Received: 4 January 2025

Revised: 20 January 2025

Accepted: 24 January 2025

Published: 28 January 2025

Citation: Yu, X.; Cui, S.; Wang, X.; Zhang, J.; Cheng, Z.; Mu, X.; Tang, B. A Multi-Level Location-Aware Approach for Session-Based News Recommendation. *Electronics* **2025**, *14*, 528. <https://doi.org/10.3390/electronics14030528>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the introduction of various news applications, there is a huge amount of news being published every day, making it difficult for users to find content of their interest.

Therefore, personalized recommendation [1] is essential, as it can provide users with a personalized news list based on their preferences.

Currently, news recommendation methods can be divided into three categories. The first is content-based news recommendation [2], which uses news text to mine user preferences and then recommends news based on the learned user preferences. The second is collaborative filtering-based news recommendation [3], which uses historical interaction information between users and news to learn implicit features of users and news, and then recommends news based on these features. The third is hybrid recommendation methods that combine content-based news recommendation and collaborative filtering-based news recommendation [4], which use both explicit and implicit features for news recommendation. Regardless of the news recommendation method used, there is a challenge when users do not register or log in to the news app, which makes it difficult to obtain their news reading history and learn their preferences. To alleviate this problem, session-based news recommendation [5] has been proposed, which learns users' short-term preferences by analyzing the news they read within a short period and recommends news.

Existing works on session-based news recommendation often use natural language processing models to process news text information, and then input the news sequence into a recurrent neural network (RNN) [6] to learn users' short-term preferences. Finally, news recommendation is made based on the similarity between user and news features. Zhang et al. [7] proposed DAINN, which uses attention mechanisms to capture user topic preferences in news text. Moreira et al. [8] proposed the session-based news recommendation framework CHAMELEON, which utilizes not only news text information but also news and user contextual information to learn users' short-term preferences for session-based news recommendation. Sheu et al. [9] proposed CAGE, which introduces a knowledge graph to learn semantic features for news and uses graph neural networks for information aggregation within the session. Although existing session-based news recommendation methods extract news text features sufficiently, they all ignore the use of location information related to news and session, which leads to suboptimal feature extraction. Moreover, these models all use BPR or BCE loss for model training and employ ranking metrics for evaluation, which may cause inconsistency between the training objective and evaluation metrics, resulting in suboptimal recommendation performance [10].

In order to alleviate the problems existing in current session-based news recommendations, such as location awareness, this paper proposes a multi-level location-aware approach for session-based news recommendation (MLA4SNR) that fully utilizes location information through multi-level location-aware, achieving an improvement in recommendation performance. MLA4SNR mainly consists of the following parts: (1) text and location feature extraction layer, which uses BERT and knowledge graph to extract news and location features, respectively; (2) news-level location-aware layer, which establishes a news-location heterogeneous graph based on the location related to news and designs a graph element-wise attention network for mining high-order relationships between news and locations; (3) session feature extraction layer, which designs a session feature extraction network based on the Transformer model; (4) session-level location-aware layer, which constructs a session-location heterogeneous graph based on the location related to the session and uses the graph attention network proposed in (2) to mine high-order relationships between sessions and locations; (5) click-through rate prediction layer, which uses a multi-layer neural network to predict the probability that a given user will click on news. In addition, we designed a loss function based on the commonly used news recommendation evaluation metric NDCG.

2. Related Work

2.1. News Recommendation

News recommendation models can be divided into traditional news recommendation models and session-based news recommendation models. For traditional news recommendation methods, as shown in Table 1, Goossen et al. [11] proposed CF-IDF, which is a concept-based approach based on TF-IDF (a common weighting technique for information retrieval and text mining) [12] and only considers the key concepts in the text, instead of using all words. Zheng et al. [13] introduced DRN, a deep learning-based online personalized news recommendation framework. Tian et al. [14] proposed KOPRA, which directly identifies entities related to user interests and derives the final user representation. Jia et al. [15] proposed RMBert, a cycle memory inference network based on Bert (a pre-training language model based on Transformer architecture). Considering that users may have different interests in each piece of news or topic of news, the model can dynamically learn the vectors of news and users and model the interaction between them. However, none of the above models mentioned or used geographical location information. Therefore, they cannot consider the current geographical location of users or the geographical relevance of news when recommending news.

Table 1. Traditional news recommendation methods.

| Model | CF-IDF | DRN | KOPRA | RMBert |
|--------|--|---|---|---|
| Author | Goossen | Zheng | Tian | Jia |
| Method | Only consider key concepts in the text | A deep learning-based online personalized news recommendation framework | Directly identify entities related to user interests and derive the final user representation | Circular memory reasoning network based on Bert |

For session-based news recommendation, as shown in Table 2, Sottocornola et al. [16] proposed a hybrid recommendation model for news sessions that completes news recommendations by mining users' short-term intentions. Hidasi et al. [6] proposed GRU4Rec, which uses GRU (a lightweight cyclic neural network variant) to extract users' short-term preferences in sessions. Zhang et al. [7] proposed DAINN, which uses a dynamic attention network to simulate users' dynamic interests. Sheu et al. [9] proposed CAGE, which enriches the semantic information of entities in news articles by constructing an auxiliary knowledge graph and enhances article embedding using graph convolutional networks. Existing session-based news recommendation methods focus on extracting news text features but ignore the use of user and news geographic location information.

Table 2. Conversation-based news recommendation method.

| Model | Mixed Recommendation Model | GRU4Rec | DAINN | CAGE |
|--------|---|---|---|---|
| Author | Sottocornola | Hidasi | Zhang | Sheu |
| Method | Digging into users' short-term intentions | Using GRU to extract users' short-term preferences during a session | Using dynamic attention networks to simulate users' dynamic interests | Building auxiliary knowledge graphs enriches the semantic information of entities in news articles, and using graph convolutional networks enhances article embedding |

2.2. Graph Neural Network

In recent years, a large amount of graph data has emerged, such as social networks, molecular structures, and knowledge graphs, which contain a wealth of information. Graph data is a data structure used to represent entities and their relationships. It is widely used to model and analyze various complex relationships and networks. The mining of information in such graph data has gradually become a research hotspot. Graph neural network models are mainly divided into frequency domain models and spatial domain models. Frequency domain models treat graph-structured data as a signal and use signal processing techniques to process them. As shown in Table 3, Shuman et al. [17] proposed the first frequency domain graph neural network model, which transformed spatial domain graph data to the frequency domain and used filtering operations to complete graph convolution. Spatial domain models directly process graph data and use weighted aggregation strategies similar to CNN to extract node features. Kipf et al. [18] proposed the first spatial domain graph neural network model, GCN, and applied it to semi-supervised classification tasks. Hamilton et al. [19] believed that when GCN aggregates information, all neighboring nodes of a target node need to participate, which is inefficient. They proposed GraphSAGE, which samples a portion of neighboring nodes to participate in information aggregation. Veličković et al. [20] introduced the attention mechanism into graph convolutional neural networks for the first time, proposing GAT. Graph neural networks have also been widely used in recommendation systems. Berg et al. [21] proposed GCMC, the first graph convolution-based matrix completion algorithm. He et al. [22] proposed LightGCN, which removed the mapping and nonlinear activation parts of the original graph convolution model, reducing computational complexity and making it more suitable for collaborative filtering recommendation tasks.

Table 3. Graph neural network model.

| Model | Frequency Domain Graph Neural Network Model | GCN | GraphSAGE | GAT |
|--------|---|-------------------------------------|--|--|
| Author | Shuman | Kipf | Hamilton | Veličković |
| Method | Converting spatial domain graph data to frequency domain and using filtering operations to complete graph convolution | Semi supervised classification task | Sampling a portion of adjacent nodes to participate in information aggregation | Introducing attention mechanism into graph convolutional neural networks |

2.3. Attention Mechanism

Attention mechanism is a neural network architecture component that simulates human visual attention. It aims to make the model dynamically focus on different parts of input data to improve processing efficiency and performance. Currently, attention mechanisms are widely used in click-through rate prediction, rating prediction, multimedia recommendation, and other scenarios, and can effectively improve recommendation performance. Mnih et al. [23] first proposed attention mechanisms, suggesting that the impact of inputs on outputs can be reflected by weights. Vaswani et al. [24] proposed a multi-head self-attention mechanism, which realizes attention calculation in multiple feature spaces. Attention mechanisms are widely used in the news recommendation field. Zhu et al. [25] proposed a method of news recommendation using a deep attention neural network, which combines the convolutional neural network, recurrent neural network and attention mechanism, fully considering the importance of the order of user clicks on news. Zhao et al. [26] believed that previous news recommendation methods mostly capture the interaction between users and news in a static way. However, news is very complex and diverse. Blindly compressing all content into one vector may affect the results of news

recommendations. To solve this problem, they proposed a dynamic news recommendation model that integrates continuous time information into attention weight calculation to better explore user’s dynamic preferences. Wu et al. [27] addressed the problem of the different interests of different users and differences in the interests of the same user and proposed a news recommendation model with personalized attention. The model consists of two parts: the news representation module and the user representation module.

3. Proposed Model

In this section, a multi-level location-aware approach for session-based news recommendation (MLA4SNR) will be proposed. Firstly, we introduce the problem formulation of this paper. Then, the main structure of the model is outlined, and finally, the details of each part of the model are introduced. The main symbols and their explanations are shown in Table 4.

Table 4. Symbols and explanations.

| Symbol | Explanation |
|---|---|
| \mathcal{V} | The news set |
| \mathcal{U} | The session set |
| \mathcal{P} | The location set |
| \mathcal{S}^i | The i -th session. |
| $p^{\mathcal{S}^i}$ | The location where session \mathcal{S}^i occurs |
| $\mathcal{V}^{\mathcal{S}^i}$ | The candidate news set for session \mathcal{S}^i |
| \mathcal{P}^v | The set of locations related to news article v |
| \mathcal{E}_v | The news text feature set |
| \mathcal{E}_p | The location feature set |
| $\mathbf{E}_{news}^{\mathcal{S}^i}$ | The news feature matrix of session \mathcal{S}^i |
| $\mathbf{E}_{position}^{\mathcal{S}^i}$ | The sequence encoding matrix of session \mathcal{S}^i |

3.1. Problem Formulation

In this paper, the scenario is an anonymous session-based news recommendation, and the news set is denoted as $\mathcal{V} = \{v_1, v_2, \dots, v_{|\mathcal{V}|}\}$, set length is $|\mathcal{V}|$, the anonymous session set is denoted as $\mathcal{U} = \{\mathcal{S}^1, \mathcal{S}^2, \dots, \mathcal{S}^{|\mathcal{U}|}\}$, with a length of $|\mathcal{U}|$, and the set of geographical locations is denoted as $\mathcal{P} = \{p_1, p_2, \dots, p_{|\mathcal{P}|}\}$, with a length of $|\mathcal{P}|$. Given an anonymous session $\mathcal{S}^i = \{s_1^i, s_2^i, \dots, s_L^i\}$, $i = 1, 2, \dots, |\mathcal{U}|$, $L = |\mathcal{S}^i|$, $s_j^i \in \mathcal{V}$ denotes the news IDs in this session, and the length of the session is $|\mathcal{S}^i|$; the location where the session occurs is $p^{\mathcal{S}^i} \in \mathcal{P}$, and the set of candidate news for this session is $\mathcal{V}^{\mathcal{S}^i} = \{v_1, v_2, \dots, v_{|\mathcal{V}-\mathcal{S}^i|}\}$, $v_j \in \mathcal{V} - \mathcal{S}^i$. Given a news article v , its related location set is $\mathcal{P}^v = \{p_1^v, p_2^v, \dots, p_{|\mathcal{P}^v|}^v\}$, $p_j^v \in \mathcal{P}$ with a length of $|\mathcal{P}^v|$. Based on the above definitions, as shown in Figure 1, the problem formulation in this paper is as follows: using the text information and related location information of the news in the given session \mathcal{S}^i , learn the short-term preferences of the anonymous user to which the session belongs, and use the short-term preferences of the anonymous user to predict the probability $\hat{y}_{\mathcal{S}^i, v_j^{\mathcal{S}^i}}$ of clicking on the candidate news $v_j^{\mathcal{S}^i}$.

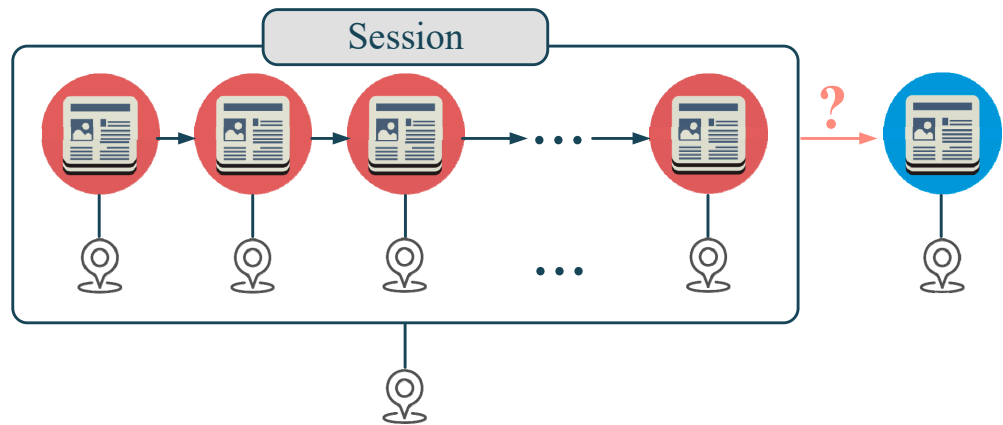


Figure 1. Problem formulation.

3.2. Model Overview

To address the insufficient use of location information in existing models and the inconsistency between their objective functions and evaluation metrics, this paper proposes a multi-level location-aware approach for session-based news recommendation (MLA4SNR). The main structure of MLA4SNR is shown in Figure 2, which consists of the following components: (1) the text and location feature extraction layer, which uses BERT and knowledge graph to extract news and location features, respectively; (2) the news-level location-aware layer, which constructs a news-location heterogeneous graph \mathcal{G}_{np} based on the location related to the news, and designs a graph element-wise attention network for mining high-order relationships between news and locations; (3) the session feature extraction layer, which designed with a Transformer-based model for extracting features from sessions; (4) the session-level location-aware layer, which constructs a session-location heterogeneous graph \mathcal{G}_{sp} based on the location relevant to the session and uses the graph attention network proposed in (2) to mine the high-order relationship between the session and the location; (5) the click-through rate prediction layer, which uses a multi-layer neural network to predict the probability of a given user clicking on recommended news. In terms of the model’s loss function, we designed a novel loss function based on the commonly used evaluation metric NDCG for news recommendation.

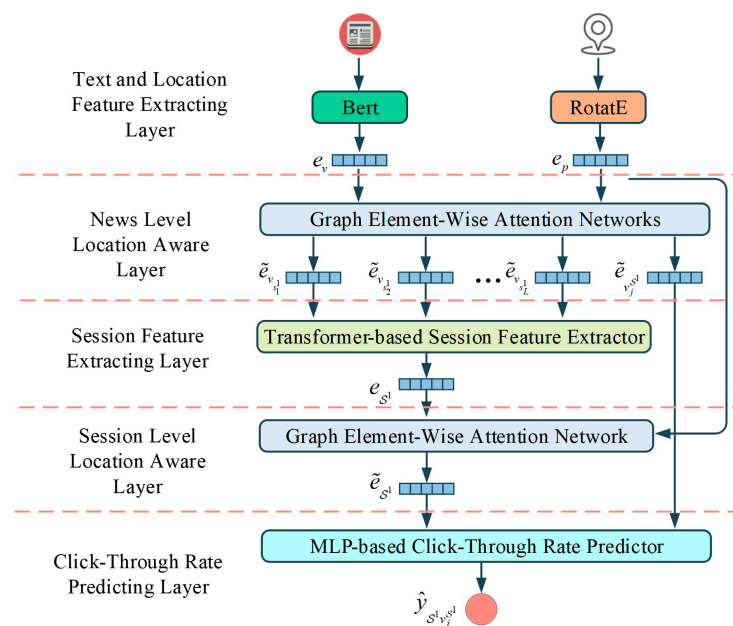


Figure 2. Diagram of the MLA4SNR structure.

3.3. Text and Location Feature Extraction Layer

3.3.1. Text Feature Extraction

The text features of a news article include the title and content, which represent the main information of the news. With the development of natural language processing technology, many high-performance natural language models have been proposed, among which the Bert model has high accuracy and efficiency compared to RNN (a neural network for processing sequence data) and CNN (a neural network for processing structured data)-based natural language models. Therefore, to extract more accurate news text features, this paper uses the Bert model for text feature extraction. In this process, first, the news title and content are integrated into a single text, then the pre-trained Bert model is fine-tuned using the entire news corpus, and finally, the fine-tuned Bert model is used to extract the news text feature set $\mathcal{E}_V = \{e_{v_1}, e_{v_2}, \dots, e_{v_{|V|}}\}$, $v_j \in \mathcal{V}, e_{v_j} \in \mathbb{R}^d$.

3.3.2. Location Feature Extraction

News is closely related to its location, and the news that users read varies depending on their location. Therefore, it is essential to focus on geographic location information and accurately represent it in session-based news recommendations. To represent each geographic location, external knowledge from the knowledge graph is introduced to learn its basic features. Specifically, first, a knowledge graph $KG = \{(h, r, t) \mid h, t \in \mathcal{E}, r \in \mathcal{R}\}$ containing geographic location entities and other related entities is constructed based on the Wikipedia knowledge graph using entity linking methods [28]. Then, the RotatE[29] algorithm based on relational reasoning is used to extract entity features from the knowledge graph. The geographic location feature set is denoted by $\mathcal{E}_P = \{e_{p_1}, e_{p_2}, \dots, e_{p_{|P|}}\}$, $p_j \in \mathcal{P}, e_{p_j} \in \mathbb{R}^d$.

3.4. News-Level Location-Aware Layer

The content of a news article may be highly related to the characteristics of its associated geographical location. For example, a news article about the Harbin Ice and Snow World is highly related to the cold weather in Harbin. There is often some connection between news articles related to the same geographical location. For example, one news article may be a follow-up to another. Therefore, mining the relationship between news articles and their related locations is crucial for enhancing the representation of news. To achieve this, we constructed a heterogeneous news-location graph $\mathcal{G}_{VP} = \{(v, p) \mid v \in \mathcal{V}, p \in \mathcal{P}\}$ to represent the relationship between news and locations. The nodes represent news articles and locations, and there is an edge between them when a news article is related to a location.

Graph neural networks (GNNs) have become an important tool for mining node relationships in graphs in recent years, and have been widely applied in various fields. GNNs often enhance the relationships between nodes and the features of nodes by propagating information between nodes in the graph. In the news-location heterogeneous graph, the node relationships are as follows: given a piece of news, its relationship with locations is one-to-many, which means that a piece of news may contain multiple related locations, each with its own characteristics, and the contribution of each location to enhance news features should be different. Conversely, given a location, its relationship with news is also one-to-many, and news that is more related to the location has a greater contribution to enhancing its features. Based on the above relationships, in order to fully explore the high-order relationships between news and locations and learn the degree of correlation between nodes, this paper designs a graph element-wise attention network

(GEAN) using element-wise attention mechanism to precisely learn the weights between nodes and their neighboring nodes, and complete the element-wise information propagation to extract more accurate news features.

Given a node $u_i \in \mathcal{V} \cup \mathcal{P}$ in the news-location graph, if the node is a news node, its initial feature is $e_{u_i} \in \mathcal{E}_V$. If the node is a location node, its initial feature is $e_{u_i} \in \mathcal{E}_P$. In layer $l+1$, the process of information propagation from neighboring nodes is divided into three steps. The first step is to use the element-wise attention mechanism to calculate the element-wise importance of neighboring nodes, that is, the importance of each position in the feature vector of neighboring nodes to node u_i . The second step is to aggregate the information propagated by each neighboring node using the element-wise weights calculated in the first step. The third step is to update the node feature using the propagated information. These three steps can be summarized by the following formula:

$$e_{u_i}^{(l+1)} = \sum_{u_j \in \mathcal{N}_{u_i}} \alpha_{u_i u_j}^l \odot \mathbf{W}^l e_{u_j}^l \tag{1}$$

In the equation, \mathcal{N}_{u_i} represents the neighbor nodes of node u_i , $\alpha_{u_i u_j}^l \in \mathbb{R}^d$ is the weight vector of element-wise attention, \odot represents element-wise multiplication, \mathbf{W}^l is the weight matrix of the l layer. The calculation process of $\alpha_{u_i u_j}^l$ is as follows: firstly, a single-layer fully connected network is used to calculate the bit-level relevance score vector $c_{u_i u_j}^l \in \mathbb{R}^d$ between node u_i and node u_j , which is calculated as follows:

$$c_{u_i u_j}^l = \text{relu}(\mathbf{W}_{att}^l [e_{u_i}^l \oplus e_{u_j}^l]) \tag{2}$$

where relu is a nonlinear activation function, \mathbf{W}_{att}^l is the weight matrix, \oplus is the concatenation of vectors, and then the following formula is used to normalize the bit-level relevance score to obtain the bit-level attention weight vector,

$$\left(\alpha_{u_i u_j}^l\right)_k = \frac{\exp\left(\left(c_{u_i u_j}^l\right)_k\right)}{\sum_{u_h \in \mathcal{N}_{u_i}} \exp\left(\left(c_{u_i u_h}^l\right)_k\right)} \tag{3}$$

where $k=1,2,\dots,d$ represents the k -th position of the graph element-wise attention network's attention weight vector.

After information propagation through L layers of graph element-wise attention networks, the final feature \tilde{e}_{u_i} of node u_i is obtained by weighted summation of node features across all layers, and the calculation formula is as follows:

$$\tilde{e}_{u_i} = e_{u_i} + \sum_{l=1}^L \frac{1}{l+1} e_{u_i}^{(l)} \tag{4}$$

where $\frac{1}{l+1}$ is a weight coefficient used to prevent excessive smoothing of node features.

Using graph attention networks as shown in Figure 3 on the news-location heterogeneous graph, relationships between news and locations can be explored to obtain an enhanced set of news features $\tilde{\mathcal{E}}_V = \{\tilde{e}_{v_1}, \tilde{e}_{v_2}, \dots, \tilde{e}_{v_{|\mathcal{V}|}}\}$, $v_j \in \mathcal{V}, \tilde{e}_{v_j} \in \mathbb{R}^d$.

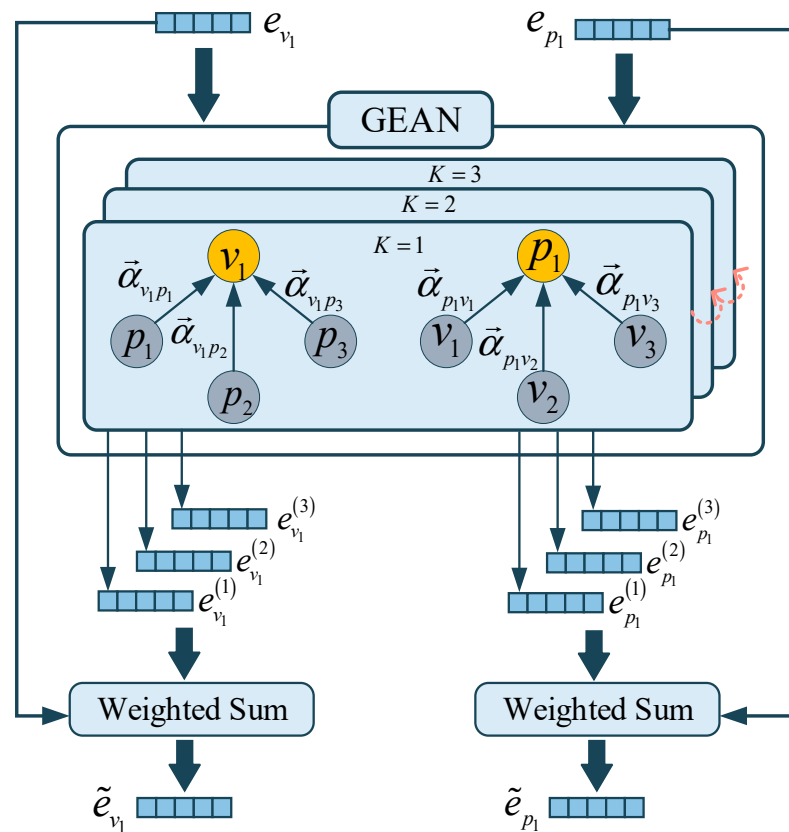


Figure 3. Location-aware news level.

3.5. Session Feature Extraction Layer

In the context of session-based recommendations, users are anonymous, and the extraction of session features can only rely on the limited news in the session. Therefore, it is critical to extract session features accurately and efficiently from a limited sequence of news in the session. Transformer models have been widely used for sequence data processing and outperform recurrent neural networks (RNNs) and convolutional neural networks (CNNs) in terms of both accuracy and efficiency. To achieve accurate feature extraction for sessions, this paper designs a Transformer-based session feature extraction network, which can extract session features accurately and efficiently. The main structure of this network is shown in Figure 4 and consists of an input layer, a feature extraction layer, and an output layer. The input layer is responsible for constructing the input of the feature extraction layer using enhanced news features, and the feature extraction layer is responsible for mining the relationships between news in the session from the perspective of features and sequences, and outputting the session features. The output layer is responsible for transforming the session feature matrix into a fixed-length feature vector. The following sections will provide a detailed description of these two parts.

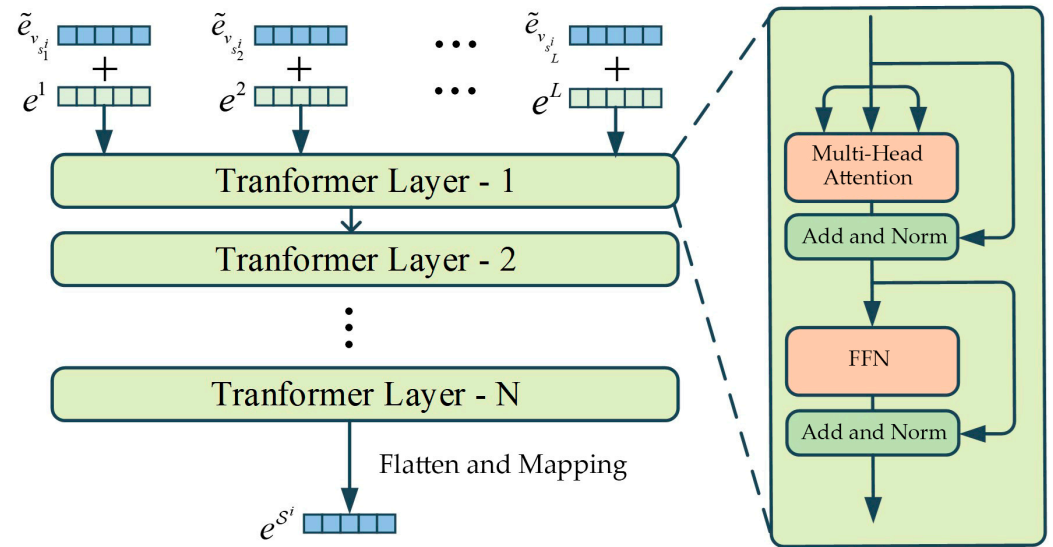


Figure 4. Transformer-based session feature extraction network.

3.5.1. Input Layer

The input layer combines a set of news features related to a session with positional encoding to obtain the corresponding feature matrix for the session. Specifically, given a session $S^i = \{s_1^i, s_2^i, \dots, s_L^i\}$ composed of L news, the news features are first combined to form a news feature matrix $\mathbf{E}_{news}^{S^i} = [\tilde{e}_{v_{s_1}^i}, \tilde{e}_{v_{s_2}^i}, \dots, \tilde{e}_{v_{s_L}^i}]^T \in \mathbb{R}^{L \times d}$. Then, the sequence is encoded using the following formula:

$$\begin{aligned} e_{2j}^{pos} &= \sin(pos / 10000^{2j/d}) \\ e_{2j+1}^{pos} &= \cos(pos / 10000^{2j/d}) \end{aligned} \tag{5}$$

where pos represents the sequence position index and j represents the position of elements in the positional vector. The sequence encoding matrix of the session S^i is $\mathbf{E}_{position}^{S^i} = [e^1, e^2, \dots, e^L]^T \in \mathbb{R}^{L \times d}$. Finally, the news feature matrix and the sequence encoding matrix are added element-wise to obtain the feature matrix E^{S^i} of the session, which is then used as the input to the Transformer layer,

$$\mathbf{E}^{S^i} = \mathbf{E}_{news}^{S^i} + \mathbf{E}_{position}^{S^i} \tag{6}$$

3.5.2. Feature Extraction Layer

In the feature extraction layer, multiple Transformer layers are used to efficiently extract more accurate session features. As shown in Figure 4, each Transformer layer contains a multi-head attention layer and a position-wise feed-forward network.

The formula for the self-attention unit in the multi-head attention layer is as follows:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d}}\right)\mathbf{V} \tag{7}$$

where \mathbf{Q} , \mathbf{K} and \mathbf{V} are the query, key, and value matrices, respectively, all derived from the news feature matrix, and $\frac{1}{\sqrt{d}}$ is a scaling factor used to avoid gradient vanishing. In the multi-head attention layer, firstly, h self-attention units are used to process the news feature matrix in parallel in multiple feature spaces. Then, the resulting feature matrices from each attention head are concatenated, and finally, the concatenated feature

matrix is mapped back to the original feature space, with the calculation process as follows,

$$\begin{aligned} \text{head}_j &= \text{Attention} \left(\mathbf{E}^{\mathcal{S}^i} \mathbf{W}_j^Q, \mathbf{E}^{\mathcal{S}^i} \mathbf{W}_j^K, \mathbf{E}^{\mathcal{S}^i} \mathbf{W}_j^V \right) \\ \text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) &= [\text{head}_1; \dots; \text{head}_h] \mathbf{W}^O \end{aligned} \tag{8}$$

where $\mathbf{W}_j^Q, \mathbf{W}_j^K, \mathbf{W}_j^V$ and \mathbf{W}^O are weight matrices.

The feed-forward network (FFN) at the positional encoding layer is a fully connected network used to capture the non-linear relationship between different features and positions. Its calculation formula is as follows:

$$FFN(x) = \text{relu}(x\mathbf{W}_1 + b_1)\mathbf{W}_2 + b_2 \tag{9}$$

where \mathbf{W}_1 and \mathbf{W}_2 are weight matrices and b_1 and b_2 are bias.

Meanwhile, residual connections and layer normalization are set in both the multi-head attention layer and the feed-forward network, the former ensuring the depth of the model and reducing the difficulty of training and the latter making the training of the model more stable and efficient.

3.5.3. Output Layer

After processing through the feature extraction layer, we obtain the feature matrix $\tilde{\mathbf{E}}^{\mathcal{S}^i}$ of the session \mathcal{S}^i . To facilitate the subsequent calculations of the model, we set an output layer to convert the feature matrix extracted by the feature extraction layer into a fixed-length feature vector. During this process, first, the feature matrix is flattened row-wise to obtain the feature vector $e^{\mathcal{S}^i} = [\tilde{\mathbf{E}}_{1,:}^{\mathcal{S}^i} \oplus \tilde{\mathbf{E}}_{2,:}^{\mathcal{S}^i} \oplus \dots \oplus \tilde{\mathbf{E}}_{L,:}^{\mathcal{S}^i}]$ of the session \mathcal{S}^i . Then, a single-layer fully connected network is used to map it to a fixed dimension to obtain the session feature $e^{\mathcal{S}^i} \in \mathbb{R}^d$ for session-level position awareness.

3.6. Session-Level Location-Aware Layer

User reading preferences may be closely related to their location, for example, users may be more interested in news that is closely related to their current location. Users in the same region may also have similar reading preferences; for example, users in the same region may all be interested in news related to local customs. Therefore, it is crucial to fully explore the relationship between sessions and locations. In this paper, we use a session-location heterogeneous graph, \mathcal{G}_{vp} , to represent the relationship between sessions and locations, where there is an edge between a session and the location at which it occurs. However, as there is only one location associated with each session, each session has only one edge connected in the graph, which severely hinders information propagation in the graph. To enrich the edges related to sessions, this paper considers that if there are more than half of the news articles overlapped between two sessions, then the two sessions are considered as similar sessions, and the corresponding users may have similar preferences. An edge will be established between these two sessions in the session-position heterogeneous graph. As shown in Figure 5, similar to the position-aware news level, we use a graph-level attention network on the session-position heterogeneous graph to achieve a position-aware session level and obtain an enhanced set of session features $\tilde{\mathcal{E}}_U = \{\tilde{e}_{s_1}, \tilde{e}_{s_2}, \dots, \tilde{e}_{s_{|U|}}\}, \mathcal{S}_j \in \mathcal{U}, \tilde{e}_{s_j} \in \mathbb{R}^d$.

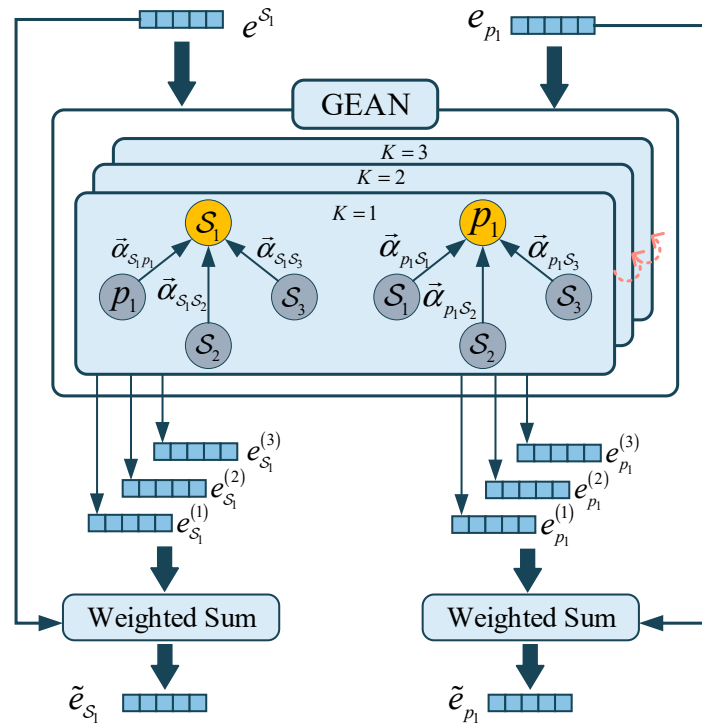


Figure 5. Location-aware session level.

3.7. Click-Through Rate Prediction Layer

To achieve better click-through rate prediction performance, we designed a neural network-based click-through rate prediction layer consisting of two fully connected layers. The input is the session features and candidate news features, and the output is the probability that the user clicks on the candidate news.

Given a session \mathcal{S}^i and a candidate news article $v_j^{\mathcal{S}^i}$, the session feature is the enhanced session feature $\tilde{e}_{\mathcal{S}^i}$ output from the session-level position-aware layer, and the candidate news feature is the enhanced news feature $\tilde{e}_{v_j^{\mathcal{S}^i}}$ output from the news-level position-aware layer. The input feature $e_{\mathcal{S}^i, v_j^{\mathcal{S}^i}} = [\tilde{e}_{\mathcal{S}^i} \oplus \tilde{e}_{v_j^{\mathcal{S}^i}}] \in \mathbb{R}^{2d}$ of the click-through rate prediction layer is obtained by concatenating the session feature and candidate news feature. The calculation formula of the click-through rate prediction layer is as follows,

$$\begin{aligned}
 f^{(0)} &= e_{\mathcal{S}^i, v_j^{\mathcal{S}^i}} \\
 f^{(1)} &= \text{relu}(f^{(0)}\mathbf{W}^{(1)} + b^{(1)}) \\
 f^{(2)} &= \text{sigmoid}(f^{(1)}\mathbf{W}^{(2)} + b^{(2)}) \\
 \hat{y}_{\mathcal{S}^i, v_j^{\mathcal{S}^i}} &= f^{(2)}
 \end{aligned} \tag{10}$$

where $\mathbf{W}^{(1)}, \mathbf{W}^{(2)}$ are weight matrices, $b^{(1)}, b^{(2)}$ are bias, $\text{relu}(\cdot), \text{sigmoid}(\cdot)$ nonlinear activation function, $\hat{y}_{\mathcal{S}^i, v_j^{\mathcal{S}^i}}$ is the predicted click-through rate.

3.8. Model Training

The performance of a model is often closely related to the quality of its loss function. A high-quality loss function can often make the model converge quickly and have good however generalization performance. Existing session-based news recommendation models often use BCE (binary cross-entropy) or BPR (Bayesian personalized ranking) as

the loss function. BCE is used to measure the difference between the prediction probability of the model and the real label. BPR optimizes the model by maximizing the probability that users prefer positive samples over negative samples. However, there is an inconsistency between these loss functions and evaluation metrics HR (hit rate) and NDCG (normalized discounted cumulative gain) [10]. To alleviate the inconsistency between the loss function and evaluation metrics, this paper proposes a loss function for the ranking metric NDCG based on a neural network ranking model [30].

For a given session \mathcal{S}^i with its corresponding user-clicked news $v_j^{\mathcal{S}^i}$, we sample $L-1$ candidate news from the candidate news set via negative sampling to construct a recommendation list of length L $[v_j^{\mathcal{S}^i}, v_1^{\mathcal{S}^i}, v_2^{\mathcal{S}^i}, \dots, v_{L-1}^{\mathcal{S}^i}]$. The clicked news $v_j^{\mathcal{S}^i}$ is placed at the top of the list followed by randomly shuffled samples. We use MLA4SNR to predict the click-through rate of each candidate news in the list, resulting in a click-through rate vector $\hat{y}_{\mathcal{S}^i} = [\hat{y}_{\mathcal{S}^i, v_j^{\mathcal{S}^i}}, \hat{y}_{\mathcal{S}^i, v_1^{\mathcal{S}^i}}, \hat{y}_{\mathcal{S}^i, v_2^{\mathcal{S}^i}}, \dots, \hat{y}_{\mathcal{S}^i, v_{L-1}^{\mathcal{S}^i}}]^T$.

Given a recommendation list of length L , the calculation process of NDCG@K is as follows:

$$NDCG@K = \sum_{k=1}^L \frac{2^{y_k} - 1}{\log_2(k + 1)} \tag{11}$$

where y_k represents the true click-through rate of the news at position k in the recommended list. The calculation of NDCG requires sorting the recommended list by click-through rate. However, the sorting process is a non-differentiable operation, which makes it difficult to optimize the model using gradient descent directly. To address this issue, we use the NeuralSort model for listwise ranking. First, we calculate the pairwise absolute difference matrix $\mathbf{A}_{\mathcal{S}^i}[k, t] = |\hat{y}_{\mathcal{S}^i, k} - \hat{y}_{\mathcal{S}^i, t}|, k = 1, 2, \dots, L, t = 1, 2, \dots, L$ and then use the following formula to compute the k -th row of the ranking matrix $\mathbf{R}_{\mathcal{S}^i}$,

$$\mathbf{R}_{\mathcal{S}^i}[k, :](\tau) = \text{soft max} \left[\left((L + 1 - 2k)\hat{y}_{\mathcal{S}^i} - \mathbf{A}_{\mathcal{S}^i} \mathbf{1} \right) / \tau \right] \tag{12}$$

where τ is the temperature coefficient, $\mathbf{1}$ is a column vector with all values equal to 1. The k -th row of the ranking matrix represents the one-hot representation of the position of the news item at position k in the list of samples $[v_j^{\mathcal{S}^i}, v_1^{\mathcal{S}^i}, v_2^{\mathcal{S}^i}, \dots, v_{L-1}^{\mathcal{S}^i}]$. For example, if a news item is in position 3 in a list of samples with length 5 and is ranked second in the final ranking, then the second row of the ranking matrix would be "00100". Finally, we obtain the loss function based on NDCG,

$$L = \frac{1}{|\mathcal{Y}|} \sum_{\mathcal{S}^i \in \mathcal{Y}} \sum_{k=1}^K \frac{2^{\mathbf{R}_{\mathcal{S}^i, k} y_{\mathcal{S}^i}} - 1}{\log_2(k + 1)} + \lambda \|\theta\|^2 \tag{13}$$

where \mathcal{Y} represents the set of samples, K represents the final length of the recommendation list, $\mathbf{R}_{\mathcal{S}^i, k}$ represents the row vector composed of the elements in the k -th row of the sorting matrix, λ is the coefficient of the regularization term, $\|\cdot\|$ represents the Frobenius norm of the matrix, and θ represents the model parameters. Based on the above loss function, we use the Adam optimizer to optimize the model parameters, which is a stochastic gradient descent-based optimizer and can make the model converge quickly at a larger learning rate.

4. Experiment

This chapter will validate and analyze the performance of the multi-level position-aware method proposed in this paper for session-based news recommendation through experiments. First, the basic settings of the experiment are introduced in Section 4.1. Then, in Section 4.2, the comparison results between MLA4SNR and baseline models are presented, and the results are analyzed. Finally, in Section 4.3, ablation and effectiveness analysis on MLA4SNR are conducted, the impact of key parameters on the model performance is explored, and the effectiveness of the proposed loss function is verified. Regarding hardware configuration, we used 500 servers and high-performance multi-core CPUs, with 64 GB of memory per server, SSD hard drive type, 88TB storage capacity, and 10 Gbps network.

4.1. Experiment Setting

4.1.1. Dataset

The dataset used in the experiment is the real news dataset Adressa, which was collected by the Norwegian University of Science and Technology and Adressavisen (a local newspaper in Trondheim, Norway). In this study, we used two versions of the dataset, one containing one week of news data and the other containing ten weeks of news data. For simplicity, the first 24 days of data from the 10-week dataset were extracted as the experimental data. News reading records were segmented into sessions using a time interval of 30 min between news within the same session, and sessions with a length greater than 20 or less than 3 were removed. For any session, the last news item in the session is considered the label or positive sample, and the remaining part is used as input to the model. For the 7-day dataset, we treated it as a fold and used the data from the first 6 days for training and the last day for testing. For the 24-day dataset, to investigate the impact of different training days on the model, it was divided into 6, 4, and 2 folds using 4, 6, and 12 days, respectively. For each fold, the data from the first 3, 5, and 7 days were used for training, and the data from the last day were used for testing. The above processing results in four data subsets: 7 days, 24 days (2 folds), 24 days (4 folds), and 24 days (6 folds). For each session, negative samples were randomly selected from the set of news items published within 300 min of the positive sample. In the training set, 5 negative samples were sampled for each session, and in the testing set, 99 negative samples were sampled for each session. The statistics of the two datasets were shown in Table 5. Additionally, for the construction of the knowledge graph, the open-source knowledge graph Wikidata, which is a knowledge base based on Wikipedia, was used. The 1-hop neighbors of geographical location entities and their corresponding relations were extracted to construct the geographical location knowledge graph. The statistics of the geographical location knowledge graph are shown in Table 6.

Table 5. Table of data statistics.

| | Number of Sessions | Number of News | Average Length of Session |
|---------|--------------------|----------------|---------------------------|
| 7 days | 460,633 | 4597 | 2.82 |
| 24 days | 1,575,240 | 16,678 | 3.21 |

Table 6. Knowledge graph dataset statistics table.

| Dataset | Number of Geographical Entities | Number of Other Entities | Number of Relationships | Number of Triples |
|---------|---------------------------------|--------------------------|-------------------------|-------------------|
| 7 days | 1687 | 45,233 | 502 | 81,414 |
| 24 days | 2547 | 69,745 | 684 | 123,584 |

4.1.2. Evaluation Metrics

To evaluate the performance of our model in real-world scenarios, a commonly used evaluation method [31] was employed, which trains the model using data from the previous few days and tests it using the remaining data. First, each dataset was divided into several folds, and then each fold was tested using the above evaluation method. Finally, the average of all fold test results was taken as the final test result. HR and NDCG were used as evaluation metrics (@10, NDCG@10). HR and NDCG were used as the evaluation metrics for the model [32]. For a session, HR is used to measure the proportion of positive samples in the recommended list among all positive samples. Since each session has only one positive sample, HR is either 0 or 1. NDCG is used to measure the quality of ranking. The HR and NDCG for each session in the validation set and the test set were calculated, and their average was taken as the final result. In addition, to evaluate the novelty and diversity of the model's recommendation results, we also introduced expected self-information with rank and relevance-sensitivity (ESI-RR) [8] and expected intra-list diversity with rank and relevance-sensitivity (EILD-RR) [8] to evaluate the model.

4.1.3. Comparison Methods and Parameter Settings

MLA4SNR was compared with five other models, including a clustering-based model, a matrix factorization-based model, and three deep learning-based models. On the dataset used in this paper, a grid search was used to tune the parameters of each model, with NDCG@10 as the metric. Detailed information regarding these models is shown below:

- Item KNN [33] is a neighbor-based approach that returns the most similar items based on the cosine similarity between the last news in a session and the rest of the news. In the experiment, the news text feature was used to calculate the cosine similarity. The core parameter recommended list length was set to 100.
- V-skNN [33] is also a neighbor-based method, similar to Item KNN, but different in that it recommends the most suitable news by considering all the news in the entire session. In the experiment, the number of similar sessions was set to 100, 200, 100, and 50 for the four data subsets, respectively, and the number of samples was set to 500, 1000, 500, and 300.
- GRU4REC [6] is an RNN-based method that uses RNN to model user sessions and predict the probability of future events in the current session (such as item click-through rate). In the experiment, the hidden state dimensions are set to 128, 256, 128, and 64, respectively, on the four data subsets, the GRU layers are all set to 3, the dropout ratios are 0.5, 0.7, 0.4, and 0.1, respectively, and the loss function is BPR.
- Chameleon [8] is a deep learning-based method that consists of two modules, one for learning news article representations and the other for providing session-based recommendations using an RNN. In the experiments, the regularization coefficient was set to 0.0001, 0.001, 0.0001, and 0.00001 for the four data subsets, respectively. The article embedding dimension was set to 1024 for all subsets, and the RNN layers were set to 2.
- CAGE [9] is a context-aware graph embedding method that constructs an auxiliary knowledge graph to mine the semantic information of news articles and uses a graph convolutional neural network to enrich the article embeddings. In the experiment, the knowledge graph embedding dimensions were set to 100, 100, 100, and 50 for the four data subsets, and the output dimensions of the first layer GCN were set to 150, 250, 125, and 125, while the output dimensions of the second layer GCN were set to 128, 128, 128, and 64.

- MLA4SNR is proposed method in this paper, in the experiments, the news embedding dimension is set to 768, the position embedding dimension is set to 768, the number of graph convolution layers in the news-level position-aware layer is set to 4, 3, 5, and 4, respectively, the number of graph convolution layers in the session-level position-aware layer is set to 2, 2, 3, and 2, respectively, the Transformer layers are set to 4, 5, 4, and 3, respectively, the attention head numbers are set to 4, 4, 4, and 2, respectively, and the regularization coefficients are set to 0.0001, 0.0001, 0.0001, and 0.00001 for the four data subsets.

4.2. Performance Comparison

As shown in Table 7, MLA4SNR was compared with the baseline models. Based on the analysis, the following conclusions can be drawn: (1) Deep learning-based models, such as GRU4REC, Chameleon, CAGE, and MLA4SNR, outperform clustering-based models, such as Item KNN and V-sknn, because deep learning-based models can extract more accurate features and learn nonlinear relationships, leading to better fitting capability. (2) Deep learning models with the incorporation of a knowledge graph, such as CAGE and MLA4SNR, perform better than models only using news features, such as GRU4REC and Chameleon, because incorporating the knowledge graph enriches the news features significantly, which benefits the recommendation performance. (3) Our proposed model, MLA4SNR, outperforms all the compared models, because, firstly, it further considers the relationship between geographical location and news/session by leveraging graph neural network models to perform location-aware at both news-level and session-level, then, it constructs a Transformer-based session feature extraction network, which has higher accuracy than traditional RNN models. Finally, an NDCG-based loss function was designed to address the problem of inconsistent evaluation metrics and model optimization objectives, which enables the model to achieve better ranking performance.

Table 7. Comparison results

| Dataset | Metrics | Item KNN | V-sknn | GRU4REC | Chameleon | CAGE | MLA4SNR |
|----------------------|------------|----------|--------|---------|-----------|--------|---------|
| 7 days | HR@10 | 57.013 | 59.125 | 60.211 | 65.145 | 66.211 | 69.515 |
| | NDCG@10 | 26.575 | 29.451 | 31.213 | 36.642 | 36.752 | 41.229 |
| | ESI-RR@10 | 39.225 | 43.12 | 44.561 | 55.46 | 56.154 | 59.356 |
| | EILD-RR@10 | 1.458 | 1.589 | 1.623 | 1.762 | 1.811 | 2.014 |
| 24 days (6 folds) | HR@10 | 48.124 | 50.472 | 51.211 | 57.133 | 58.458 | 60.584 |
| | NDCG@10 | 20.131 | 22.25 | 23.144 | 27.545 | 27.881 | 30.997 |
| | ESI-RR@10 | 29.91 | 32.131 | 33.655 | 46.322 | 47.243 | 50.003 |
| | EILD-RR@10 | 1.146 | 1.258 | 1.301 | 1.498 | 1.578 | 1.702 |
| 24 days (4 folds) | HR@10 | 55.564 | 58.452 | 59.452 | 64.121 | 65.213 | 68.115 |
| | NDCG@10 | 24.985 | 28.143 | 30.784 | 34.222 | 35.045 | 39.025 |
| | ESI-RR@10 | 37.243 | 42.687 | 43.211 | 54.011 | 54.982 | 58.68 |
| | EILD-RR@10 | 1.387 | 1.513 | 1.562 | 1.701 | 1.767 | 1.946 |
| 24 days (2 folds) | HR@10 | 63.154 | 65.147 | 66.214 | 70.213 | 71.244 | 75.387 |
| | NDCG@10 | 30.012 | 33.021 | 34.025 | 39.415 | 40.312 | 44.001 |
| | ESI-RR@10 | 45.471 | 48.556 | 49.111 | 60.014 | 60.453 | 64.568 |
| | EILD-RR@10 | 1.698 | 1.754 | 1.787 | 1.878 | 1.898 | 2.112 |

4.3. Model Analysis

4.3.1. Ablation Analysis

In order to fully investigate the impact of each module on the recommendation performance of MLA4SNR, MLA4SNR was compared with its variant models, including MLA4SNR(-NL), MLA4SNR(-SL), and MLA4SNR(-NL&SL). MLA4SNR(-NL) removes the news-level position-aware layer and directly inputs the news text features into the session feature extraction layer. MLA4SNR(-SL) removes the session-level position-aware layer and directly inputs the session features extracted by the session feature extraction layer into the click-through rate prediction layer for click-through rate prediction. MLA4SNR(-NL&SL) removes both the news-level and session-level position-aware layers, and inputs the news text features into the session feature extraction layer to extract session features. It then inputs the session features and candidate news text features into the click-through rate prediction layer for click-through rate prediction. The parameters of the above variant models are consistent with the optimal parameters of MLA4SNR. As shown in the comparison results in Figure 6, the following conclusions can be drawn: (1) Among MLA4SNR and the three variant models, MLA4SNR(-NL&SL) without using location information has the worst performance because it only uses news text features to learn session features, resulting in inaccurate extracted session features. (2) MLA4SNR(-SL) performs better than MLA4SNR(-NL) because MLA4SNR(-SL) is a position-aware news level, enhancing news features. The use of enhanced news features to learn session features also enhances the session features to some extent. Meanwhile, candidate news features are also enhanced in the position-aware news level. MLA4SNR(-NL), on the other hand, only conducts position-aware session level, enhancing only session features but not candidate news features. This indicates that the position-aware news level is more important than the position-aware session level. (3) MLA4SNR performs better than all variant models because it conducts both position-aware news levels and position-aware session levels, fully utilizing location information to learn news and session features. As a result, extracted session features and candidate news features are more accurate.

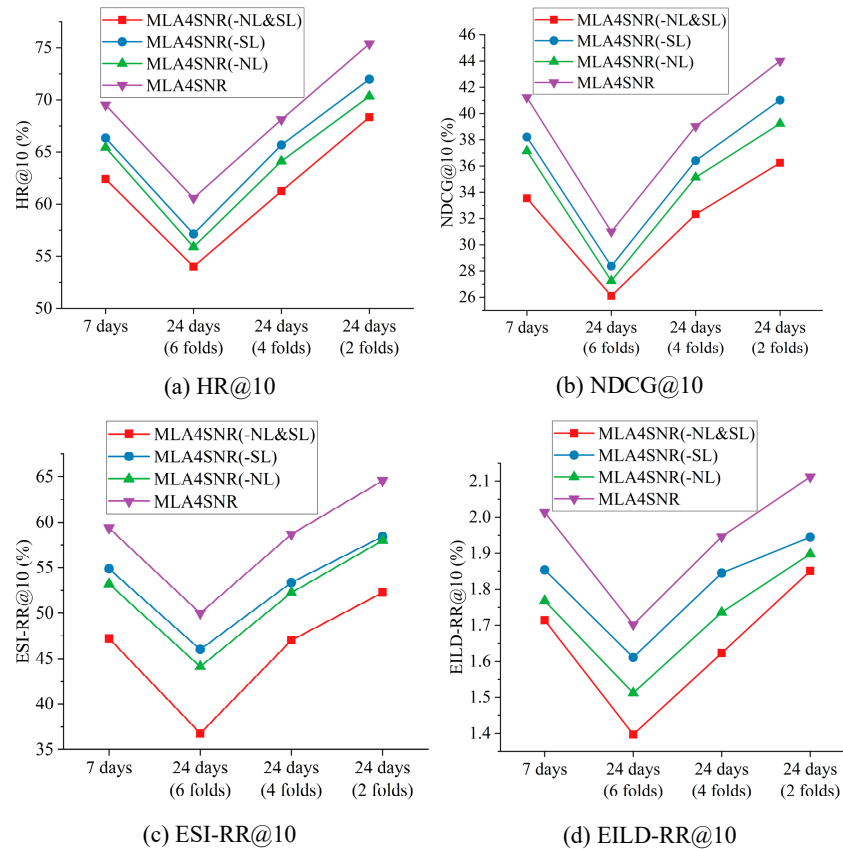


Figure 6. Ablation experiment results.

4.3.2. The Impact of the Number of Layers in Graph Element-Wise Attention Networks

In MLA4SNR, a graph element-wise attention network (GEAN) was designed to perform at the position-aware news level and session level, which enhances the news and session features. The number of layers in GEAN determines the depth of relation mining, and choosing the appropriate number of layers is crucial to achieving optimal model performance. To investigate the effect of GEAN layers on the performance of news-level and session-level position-aware layers, extensive experiments were conducted on four datasets with layer ranges from {1, 2, 3, 4, 5, 6}, using NDCG@10 as the evaluation metric. The experimental results are shown in Figure 7, where the numbers in parentheses represent the optimal number of GEAN layers in the news-level and session-level position-aware layers, respectively, along with the corresponding NDCG@10 values. It can be concluded that for the news-level position-aware layer, the optimal number of GEAN layers is around 4; for the session-level position-aware layer, the optimal number of GEAN layers is around 2, indicating that more in-depth relation mining should be performed for the news-position graph.

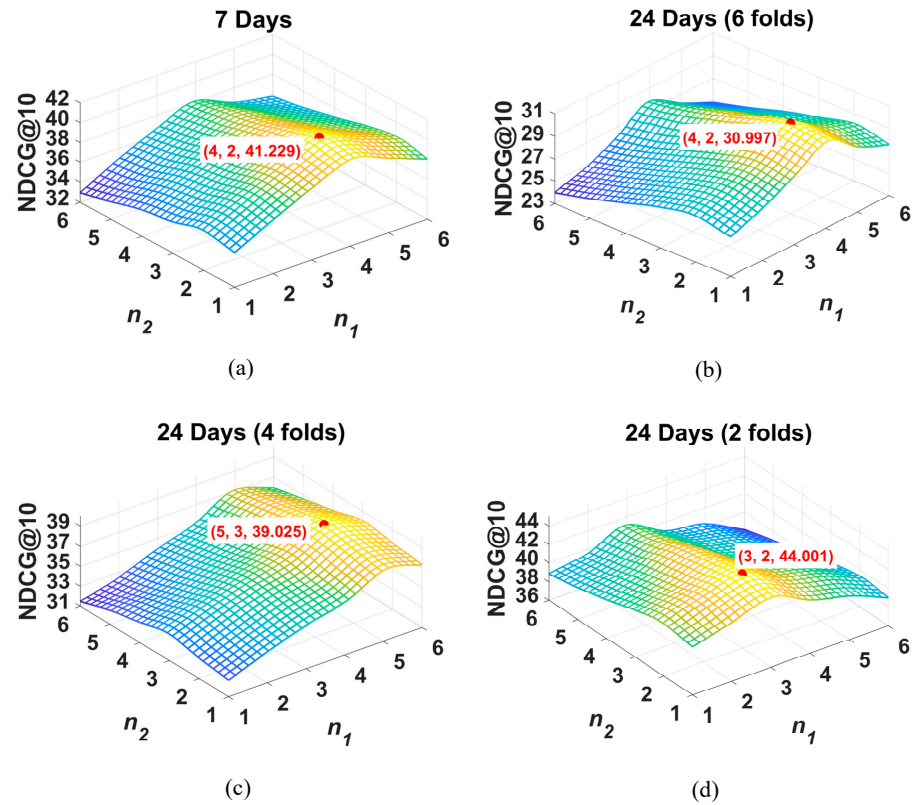


Figure 7. (a) NDCG@10 under different combinations of GEAN layers in the 7 Days dataset. (b) NDCG@10 under different combinations of GEAN layers in the 24 Days (6 folds) dataset. (c) NDCG@10 under different combinations of GEAN layers in the 24 Days (4 folds) dataset. (d) NDCG@10 under different combinations of GEAN layers in the 24 Days (2 folds) dataset.

4.3.3. The Impact of the Number of Transformer Layers and Attention Heads

To accurately extract session features, a Transformer-based session feature extraction layer, consisting of multiple Transformer layers, was designed. The number of Transformer layers and attention heads determines the performance of the session feature extraction layer. Therefore, selecting an appropriate number of layers and attention heads is crucial to extracting more accurate session features. In order to explore the impact of the number of Transformer layers and attention heads on model performance, further experiments were conducted on four datasets to investigate the performance of the model under different combinations of layers and attention heads. The range of layers was set to layers = {1,2,3,4,5}, and the range of attention heads was set to heads = {2, 4, 6, 8}, with NDCG@10 as the evaluation metric. The experimental results are shown in Figure 8, where the numbers in parentheses represent the optimal number of Transformer layers, the optimal number of attention heads, and the corresponding NDCG@10 value. On the four datasets, the optimal number of Transformer layers was around 4 layers, and the model performance decreased when the number of layers increased, as too many layers can lead to overfitting. When the number of layers decreased, the model performance also decreased, possibly because too few layers could not better fit the samples. The optimal number of attention heads was around 4, and the model performance decreased when the number of attention heads decreased, indicating that too few feature spaces are insufficient to perform high-precision attention calculations. When the number of attention heads increased, the model performance also showed varying degrees of decline because too many feature spaces would make it easier for the model to overfit.

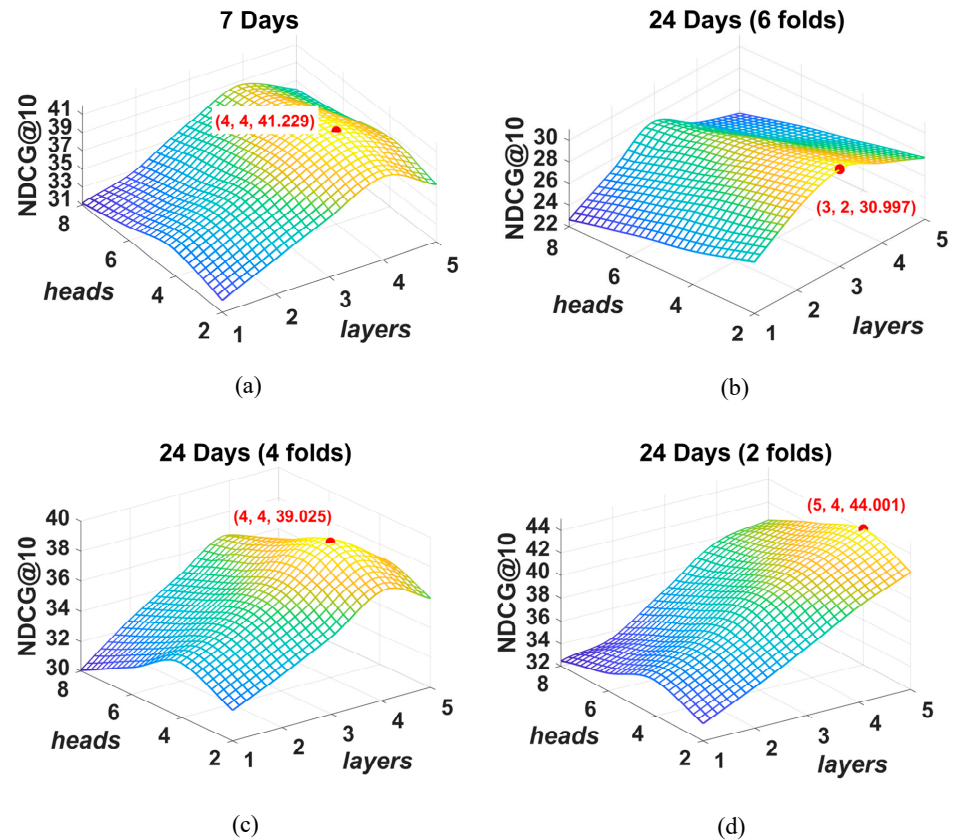


Figure 8. (a) NDCG@10 under different combinations of the number of layers and attention heads in the 7 Days dataset. (b) NDCG@10 under different combinations of the number of layers and attention heads in the 24 Days (6 folds) dataset. (c) NDCG@10 under different combinations of the number of layers and attention heads in the 24 Days (4 folds) dataset. (d) NDCG@10 under different combinations of the number of layers and attention heads in the 24 Days (2 folds) dataset.

4.3.4. Comparison of BPR, BCE, and NDCG Loss Functions

We propose a novel loss function for the ranking metric NDCG to alleviate the inconsistency between the training objective and the ranking metric, which leads to better ranking performance of the model. To validate the effectiveness of the NDCG loss function, extensive experiments were conducted on four datasets, where it was replaced with the widely used BPR and BCE losses. The evaluation metrics are two ranking metrics, HR and NDCG, and the model parameters are set to the optimal values. The experimental results in Figure 9 show that the BPR and NDCG losses designed for ranking tasks outperform the BCE loss designed for classification tasks, as the design for ranking enlarges the gap between positive and negative samples. Moreover, the proposed loss function was found to perform significantly better than the BPR loss, as the inconsistency between the training objective and the ranking metric was narrowed down by the NDCG loss.

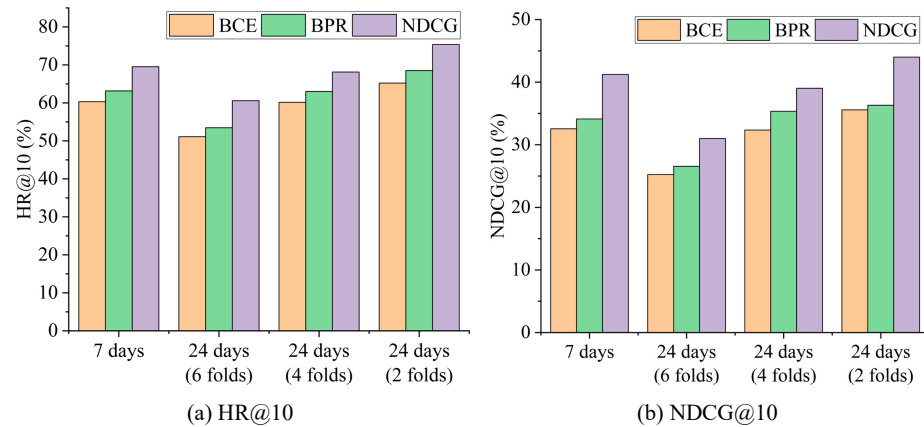


Figure 9. Performance comparison of BPR, BCE, and NDCG loss.

5. Conclusions

This paper proposes a multi-level location-aware approach for session-based news recommendation (MLA4SNR), which utilizes location-aware at both the news and session level, making full use of geographic location information, and solving the problem of existing models not fully utilizing geographic location information related to news and conversations, as well as inconsistencies between training objectives and model evaluation metrics. Firstly, a graph element-wise attention is proposed to mine high-order relationships between news and locations, achieving a location-aware news level on the news-location heterogeneous graph. Secondly, a Transformer-based session feature extraction network is used to extract session features. Thirdly, graph element-wise attention is used to mine high-order relationships between sessions and locations, achieving session-level location awareness on the session-location heterogeneous graph. Finally, a loss function designed specifically for the NDCG ranking metric is used to train the model, improving the model's ranking performance. Experimental results on a real news dataset demonstrate that MLA4SNR outperforms the state-of-the-art models. In addition, our research only focuses on improving the accuracy of model recommendations, and there are still shortcomings in terms of efficiency. We will pay attention to how to improve the efficiency of model recommendations in future research. The main contributions of this paper are as follows:

- A multi-level location-aware session-based news recommendation algorithm has been proposed, which realizes multi-level location-awareness at both the news and session levels.
- A session feature extraction network based on Transformer is designed to achieve high-precision and high-efficiency extraction of session features.
- A new ranking-based loss function is designed based on the evaluation metric NDCG, which improves the model's ranking performance.
- Experimental results on real-world news datasets demonstrate that the performance of MLA4SNR significantly outperforms the baseline.

Author Contributions: Conceptualization, X.Y.; methodology, X.Y.; software, S.C.; validation, S.C., X.W. and J.Z.; formal analysis, Z.C.; investigation, X.M.; resources, S.C.; data curation, S.C.; writing—original draft preparation, S.C.; writing—review and editing, X.Y. and Bin Tang; visualization, S.C.; supervision, X.Y. and B.T.; project administration, X.Y.; funding acquisition, X.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work is jointly sponsored by the National Natural Science Foundation of China (62172249, 62472441) and the Natural Science Foundation of Qingdao City (24-8-4-zrjj-3-ich).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Dias, M.B.; Locher, D.; Li, M.; El-Deredy, W.; Lisboa, P.J. The value of personalised recommender systems to e-business: A case study. In Proceedings of the 2008 ACM Conference on Recommender Systems, Lausanne, Switzerland, 23–25 October 2008; pp. 291–294.
2. Joseph, K.; Jiang, H. Content based news recommendation via shortest entity distance over knowledge graphs. In Proceedings of the 2019 World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 690–699.
3. Das, A.S.; Datar, M.; Garg, A.; Rajaram, S. 2007. Google news personalization: Scalable online collaborative filtering. In Proceedings of the 16th International Conference on World Wide Web, Edmonton, AB, Canada, 8–12 May 2007; pp. 271–280.
4. Wang, H.; Zhang, F.; Xie, X.; Guo, M. DKN: Deep knowledge-aware network for news recommendation. In Proceedings of the 2018 World Wide Web Conference, Lyon, France, 23–27 April 2018; pp. 1835–1844.
5. de Souza Pereira Moreira, G.; Ferreira, F.; da Cunha, A.M. News session-based recommendations using deep neural networks. In Proceedings of the 3rd Workshop on Deep Learning for Recommender Systems, Vancouver, BC, Canada, 6 October 2018; pp. 15–23.
6. Hidasi, B.; Karatzoglou, A. Recurrent neural networks with top-k gains for session-based recommendations. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Torino, Italy, 22–26 October 2018; pp. 843–852.
7. Zhang, L.; Liu, P.; Gulla, J.A. Dynamic attention-integrated neural network for session-based news recommendation. *Mach. Learn.* **2019**, *108*, 1851–1875, <https://doi.org/10.1007/s10994-018-05777-9>.
8. Moreira, G.; Jannach, D.; da Cunha, A.M. Contextual Hybrid Session-based News Recommendation with Recurrent Neural Networks. *IEEE Access* **2019**, *7*, 169185–169203.
9. Sheu, H.S.; Chu, Z.; Qi, D.; Li, S. Knowledge-Guided Article Embedding Refinement for Session-Based News Recommendation. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 7921–7927.
10. Tang, S.; Luo, F.; Wu, J. Smooth-AUC: Smoothing the Path Towards Rank-based CTR Prediction. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, 11–15 July 2022; pp. 2400–2404.
11. Goossen, F.; Ijntema, W.; Frasinca, F.; Hogenboom, F.; Kaymak, U. News personalization using the cf-idf semantic recommender. In Proceedings of the International Conference on Web Intelligence, Mining and Semantics, Sogndal, Norway, 25–27 May 2011; pp. 1–12.
12. Jones, K.S. A statistical interpretation of term specificity and its application in retrieval. *J. Doc.* **2004**, *60*, 493–502.
13. Zheng, G.; Zhang, F.; Zheng, Z.; Xiang, Y.; Yuan, N.J.; Xie, X.; Li, Z. DRN: A Deep Reinforcement Learning Framework for News Recommendation. In Proceedings of the 2018 World Wide Web Conference, Lyon, France, 23–27 April 2018.
14. Tian, Y.; Yang, Y.; Ren, X.; Wang, P.; Wu, F.; Wang, Q.; Li, C. Joint Knowledge Pruning and Recurrent Graph Convolution for News Recommendation. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, Online, 11–15 July 2021.
15. Jia, Q.; Li, J.; Zhang, Q.; He, X.; Zhu, J. RMBERT: News recommendation via recurrent reasoning memory network over BERT. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Online, 11–15 July 2021; pp. 1773–1777.
16. Sottocornola, G.; Symeonidis, P.; Zanker, M. Session-based news recommendations. In Proceedings of the Companion Proceedings of the the Web Conference 2018, Lyon, France, 23–27 April 2018; pp. 1395–1399.
17. Shuman, D.I.; Narang, S.K.; Frossard, P.; Ortega, A.; Vandergheynst, P. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Process. Mag.* **2013**, *30*, 83–98.
18. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. In Proceedings of the International Conference on Learning Representations, San Juan, Puerto Rico, 2–4 May 2016.
19. Hamilton, W.; Ying, Z.; Leskovec, J. Inductive representation learning on large graphs. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.

20. Velickovic, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. Graph Attention Networks. In Proceedings of the International Conference on Learning Representations, Toulon, France, 24–26 April 2017.
21. Berg R, Kipf T N, Welling M . Graph Convolutional Matrix Completion. *arXiv* **2017**, arXiv:1706.02263.
22. He, X.; Deng, K.; Wang, X.; Li, Y.; Zhang, Y.; Wang, M. Lightgcn: Simplifying and powering graph convolution network for recommendation. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, China, 25–30 July 2020; pp. 639–648.
23. Mnih, V.; Heess, N.; Graves, A. Recurrent models of visual attention. In Proceedings of the Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; Volume 27.
24. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhinal, I. Attention is all you need. In Proceedings of the Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.
25. Zhu, Q.; Zhou, X.; Song, Z.; Tan, J.; Guo, L. DAN: Deep Attention Neural Network for News Recommendation. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019.
26. Zhao, Q. D-HAN: Dynamic News Recommendation with Hierarchical Attention Network. *arXiv* **2021**, arXiv:2112.10085.
27. Wu, C.; Wu, F.; An, M.; Huang, J.; Huang, Y.; Xie, X. NPA: Neural News Recommendation with Personalized Attention. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019.
28. Milne, D.; Witten, I.H. Learning to link with wikipedia. In Proceedings of the 17th ACM Conference on Information and Knowledge Management, Napa Valley, CA, USA, 26–30 October 2008; pp. 509–518.
29. Sun, Z.; Deng, Z.H.; Nie, J.Y.; Tang, J. Rotate: Knowledge graph embedding by relational rotation in complex space. In Proceedings of the 7th International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019. Available online: <https://arxiv.org/abs/1902.10197> (accessed on 26 February 2019).
30. Grover, A.; Wang, E.; Zweig, A.; Ermon, S. Stochastic Optimization of Sorting Networks via Continuous Relaxations. In Proceedings of the International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019.
31. Jugovac, M.; Jannach, D.; Karimi, M. Streamingrec: A framework for benchmarking stream-based news recommenders. In Proceedings of the 12th ACM Conference on Recommender Systems, Vancouver, BC, Canada, 2–7 October 2018; pp. 269–273.
32. He, X.; Chen, T.; Kan, M.Y.; Chen, X. Trirank: Review-aware explainable recommendation by modeling aspects. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, Melbourne, Australia, 19–23 October 2015; pp. 1661–1670.
33. Ludewig, M.; Jannach, D. Evaluation of session-based recommendation algorithms. *User Model. User-Adapt. Interact.* **2018**, *28*, 331–390.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.