News Recommendation with Gated Linear Attention and Simplified Gated Linear Units

Debiao Yan[1], Jian Sun[1], and Jing Huang[2*]
[1]School of Artificial Intelligence and Big Data, Luzhou Vocational & Technical College, Luzhou 646000, China
[2]School of Medical Information and Engineering, Southwest Medical University, Luzhou 646000, China
E-mail: yandebiao@lzy.edu.cn, sj@lzy.edu.cn, huangjing_110@126.com
*Corresponding author

Keywords: news recommendation, gated linear attention, simplified gated linear units, user engagement

Received: May 21, 2024

As a novel approach, the News Recommendation Algorithm Based on Gated Linear Attention and Simplified Gated Linear Units (NRAGLS) designed to enhance the precision and personalization of digital news dissemination is introduced. At the core of NRAGLS are two advanced techniques: Gated Linear Attention (GLA) and Simplified Gated Linear Units (SGLU), supplemented by tensor normalization to adeptly process and analyze the multifaceted nature of news content and user interactions. This research systematically outlines the development and implementation of the NRAGLS model, emphasizing its capacity to address prevalent challenges within the news recommendation domain, including data sparsity, the dynamic evolution of news content, and the absence of explicit user feedback. Through rigorous experimental validation on the Microsoft News Dataset (MIND) and its variant, MIND-small, the NRAGLS model's performance was benchmarked against four existing recommendation systems. Metrics such as Area Under the Curve (AUC), Mean Reciprocal Rank (MRR), and Normalized Discounted Cumulative Gain (nDCG) were employed to evaluate each model's ability to predict user engagement accurately and to rank news articles effectively. The NRAGLS model demonstrated superior performance across all metrics, highlighting its robustness in handling the complexities of real-world news recommendation scenarios. Furthermore, this paper explores the resilience of the NRAGLS model to data sparsity and its adaptability to the temporal dynamics of user preferences, key factors in sustaining the relevance and effectiveness of recommendation systems. The findings reveal that the NRAGLS model exhibits remarkable consistency in performance, even under varying levels of data omission, and maintains or improves its recommendation accuracy over time.

Povzetek: Razvit je nov algoritem za priporočanje novic, temelječ na Gated Linear Attention (GLA) in Simplified Gated Linear Units (SGLU). Algoritem deluje dobro tudi pri pomanjkanju podatkov in hitrem spreminjanju vsebin.

1 Introduction

The transition from traditional print media to digital news consumption represents a pivotal change in the dynamics of information dissemination in our society [1]. This shift, predominantly fueled by the rapid proliferation of mobile devices and the advancement of internet technologies, has not only redefined the accessibility of news but also shows how it is consumed [2]. The ubiquitous presence of smartphones and tablets, coupled with high-speed internet connectivity, has made it possible for news to be accessed anytime and anywhere, breaking the barriers of time and space that once limited traditional media [3]. Furthermore, the advent of social media platforms and news aggregation websites has catalyzed a more profound change. These platforms, leveraging the power of the internet, have transformed news consumption from a passive, one-way communication into an interactive and engaging experience [4]. Social media, in particular, has become a significant news source for many, blending traditional news with user-generated content, and thus creating a more diverse and dynamic news environment [5]. The convenience offered by these digital platforms, however, is not without its challenges. The overwhelming volume of available content necessitates sophisticated news recommendation systems capable of filtering and personalizing content to suit individual preferences [6]. This, in turn, has sparked a surge in the development and refinement of algorithms designed to analyze and predict user interests based on their digital footprints.

The challenges confronting news recommendation systems in this digital era are multifaceted and complex. Firstly, the rapid pace at which news content is updated presents a significant hurdle. Unlike other domains where content remains relatively static, news is dynamic and continually evolving, necessitating algorithms that can adapt swiftly to the ever-changing landscape [7]. This characteristic of news content requires recommendation systems to be highly responsive and capable of incorporating the latest information with minimal delay. Secondly, the lack of explicit user ratings for news content adds another layer of complexity. In many online platforms, users' express preferences through ratings or reviews, providing clear signals for recommendation algorithms [8]. However, in the context of news consumption, such explicit feedback is often absent. Users may read an article but not provide any overt indication of their preference or disapproval, making it challenging to gauge their interests accurately. Additionally, the unique nature of news text itself poses distinct challenges. News articles, characterized by their specific writing style, structure, and content, require sophisticated natural language processing techniques for effective analysis [9]. The need to understand the nuances of political, economic, or cultural contexts and to discern the relevance of an article to a user's specific interests calls for advanced techniques in text analysis and understanding.

These challenges necessitate the development of innovative, robust, and adaptive news recommendation systems. Such systems must not only handle the dynamic nature of news content but also infer user preferences in the absence of explicit feedback and effectively process and analyze the unique characteristics of news text [10]. Addressing these challenges is crucial for the next generation of news recommendation systems, aiming to provide personalized and relevant news content to users in an ever-evolving digital landscape.

The application of deep learning in news recommendation systems has marked a paradigm shift in the approach to content personalization and user engagement [11]. The advent of deep learning technologies has enabled more sophisticated analysis of news content, providing the ability to extract complex patterns and user preferences [12]. However, this integration also brings forth new challenges. Specifically, the effective combination of news titles and body content through deep learning models remains a complex task. These models need to understand and correlate the succinct, attention-grabbing nature of titles with the detailed context provided in the body, ensuring that the essence of the news is accurately captured and recommended. Aboutorab et al. [13] proposed a deep learning meta-architecture for personalized news recommendations, focusing on sparse user profiling and dynamic preferences. Kumar et al. [14] introduced a 3-D convolutional network model, Yang et al. [15] which enhances news recommendation accuracy by using content and user interaction sequences. Xue et al. [16] propose a news recommendation model based on a multi-feature sequence transformer, improving the representation of news and user behaviors. Bangari et al. [17] reviewed reinforcement learning algorithms in news recommendation systems, focusing on addressing dynamic user preferences. Dellal-Hedjazi et al. [18] explored a deep learning-based recommendation system that combines demographic and content-based filtering. Liu et al. conducted a survey on recommendation systems based on deep learning, analyzing their advantages over traditional systems. Despite the impressive advancements brought about by deep learning in personalizing news recommendations, addressing the diversity in user interests remains a formidable challenge, those models often struggle with the vast range of topics and rapidly shifting interests of users. The necessity for these systems to be both dynamic and adaptable cannot be overstated. Deep learning models must be designed to continuously evolve and cater to the multifaceted profiles of users, ensuring that the recommendations are not only relevant but also sufficiently diverse. This ensures that users are engaged with a wide spectrum of topics, reflecting the richness and variety of their interests. To provide a clearer comparison of existing research, the key features, datasets used, performance metrics, and specific challenges addressed by the critical related works in news recommendation systems are summarized in Table 1.

Table 1: Summary of related works in news recommendation systems

Author	Methodology	Dataset	Key Metrics	Specific Challenges Addressed
Feng et al. [1]	Survey of various approaches	Multiple	N/A	A comprehensive review of challenges
Li and Wang [4]	Survey of personalization techniques	Multiple	N/A	User profiling, content analysis
Cheng et al. [24]	DLRASN: Deep learning with social network data	Proprietary dataset	AUC: 0.701, MRR: 0.341	Data sparsity, social influence
Wu et al. [25]	DLRec: Deep learning for time heterogeneous feedback	Adressa 1G	AUC: 0.715, nDCG: 0.413	Temporal dynamics, cold start
Liu et al. [26]	HybCF: Hybrid collaborative filtering with news hotness	Custom dataset	Precision: 0.710, Recall: 0.692	Content freshness, user interest drift
Zhang et al. [27]	DCDN: Deep cross network with relevant article features	MIND	AUC: 0.718, MRR: 0.356	Feature interaction, diversity

In light of the identified shortcomings in existing news recommendation algorithms, particularly their inadequacy in mining both short-term and long-term user interests from news content, this paper introduces an innovative approach. Our proposed method, the News Recommendation Algorithm Based on Gated Linear Attention and Simplified Gated Linear Units (NRAGLS) [19], incorporates advanced techniques such as Gated Linear Attention (GLA) and Simplified Gated Linear Units (SGLU) [20], along with tensor normalization, to construct a more effective news recommendation model. The key components of the method are as follows：

(1) The model leverages sophisticated topic modeling techniques to deeply analyze news content. By

extracting latent topics, it gains a comprehensive understanding of the thematic structures within the data, which forms the backbone of our news recommendation system.

(2) Utilizing GLA and SGLU, the model effectively integrates the diverse textual features of news, such as titles, summaries, and body content. This process results in a cohesive and precise representation of news articles, capturing their essence with greater accuracy.

(3) At the heart of our model lies its ability to dynamically adapt to both short-term and long-term shifts in user interest patterns. By combining GLA and SGLU structures, the model offers nuanced insights into user preferences. Coupled with an advanced attention mechanism and tensor normalization techniques, it delivers highly personalized news recommendations, balancing relevance and diversity to cater to the evolving tastes of users.

(4) The experimental analysis on the Microsoft News Dataset (MIND) demonstrates that the NRAGLS model outperforms existing methods in terms of recommendation accuracy and user engagement. This success highlights the model's effectiveness in handling real-world news recommendation scenarios, offering a significant advancement over traditional approaches.

2 Methods

2.1 Problem description

The primary goal of this research is to accurately predict a user's likelihood of clicking on a previously unseen news item. This involves analyzing various characteristics of the news content, including titles, categories, summaries, and thematic elements. To systematically approach this challenge, specific symbols are introduced to represent key components of our model, as outlined in Table 2.

Table 2: Symbol definitions

Symbol	Description
U	Set of users, with each user denoted as $u \in U$
N	Set of news items, with each item denoted as $n \in N$
T_n	Title of the news item n
C_n	Category of the news item n
S_n	Summary of the news item n
Θ_n	Set of themes or topics extracted from the content of news item n

The objective is to model the probability $P(click_u|n)$, representing the likelihood that a user u will click on a news item n. This probability is derived by analyzing the features T_n, C_n, S_n, and Θ for each news item n. To effectively predict user engagement with unseen news articles, our methodology involves the extraction and in-depth analysis of key features from each news item, The feature extraction process encompasses several aspects. For the title feature T_n, we employ text analysis techniques to extract keywords and phrases from news titles, capturing the core information of the news content. The category feature C_n involves encoding the news category into a numerical format, facilitating algorithmic processing and classification. Similarly, the summary feature S_n is derived by analyzing the text of news summaries to extract key information, aiding in understanding the main content of the news. Additionally, the topic feature Θ_n is obtained through Latent Dirichlet Allocation (LDA) methods [21], extracting thematic distributions from the news body to grasp deeper thematic structures.

In the feature analysis phase, we integrate these extracted features—titles, categories, summaries, and topics into a composite feature vector. This comprehensive approach lays the groundwork for subsequent user interest prediction. The composite feature vector undergoes a weighting process, where each feature is assigned, a weight based on its significance in predicting user click behavior. Moreover, we incorporate context analysis by considering factors such as the timing and geographical location of news publications and assessing their potential impact on user interests. Through these multifaceted methods, we aim to analyze news content from various perspectives, thereby establishing a solid foundation for the development of accurate user interest models and recommendation algorithms.

2.2 Model design

Our method integrates several advanced technologies, including GLA, SGLU, and tensor normalization, to construct a more effective news recommendation model. The essence of this approach lies in its ability not only to process multi-dimensional features of news content but also to capture both short-term and long-term user interests, thereby offering more accurate and personalized news recommendations. In the news representation module of our algorithm, we amalgamate multiple features of news such as titles, categories, summaries, and topics to learn a unified representation vector of news. This design allows for a more comprehensive description of the news content and attributes. By combining and encoding these features, our model acquires rich and precise information, significantly enhancing its performance. The user representation module focuses on capturing the temporal interest features of users through the GLA and SGLU networks, integrating both long-term and short-term interest features to establish a user representation vector. Initially, historical click-through data of users is transformed into representative news vectors to better reflect the intrinsic attributes of the news. Subsequently, a structured network is employed for continuous analysis of user interests, adaptively extracting news features that are more relevant to user preferences. The final stage of our model involves the integration of long-term and short-term interest features, forming a comprehensive interest vector for each user. The calculation of the click-through probability for candidate news items is performed using the inner product of this user vector with the news vectors, effectively enhancing the model's predictive accuracy. A distinctive aspect of our approach is its effective combination of users' long-term and short-term interests, which significantly improves the accuracy of predicting user preferences. The architecture and operational flow of our NRAGLS model are illustrated in Figure 1, providing a visual representation of the integrative process encompassing both the news and user representation modules, along with the mechanisms for predictive performance enhancement.

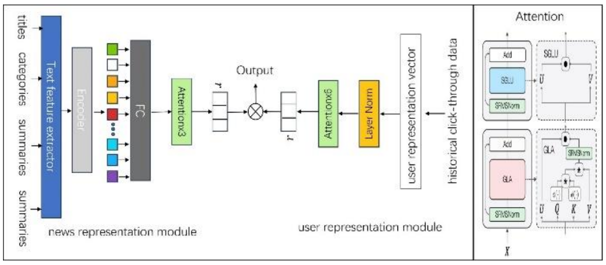

Figure 1: Architecture of the NRAGLS news recommendation model

In light of the comprehensive methodology outlined for predicting user click behavior, the integration of GLA and SGLU within the NRAGLS model serves as a pivotal enhancement to address the nuanced task of capturing user interests. The symbols defined in Table 2 represent the multifaceted features of news content which are crucial for predicting user interactions with news items. The advanced technologies of GLA and SGLU are leveraged to process these features systematically and dynamically. Given the symbols from Table 2, it defines the mathematical operations of the NRAGLS model as follows: For each news item n with title T_n, category C_n, sum thematic elements Θ_n, we extract feature vectors using the following equations:

$$\begin{aligned} \mathbf{V}_{T_n} &= \text{Embedding }(T_n) \\ \mathbf{V}_{C_n} &= \text{EncodeCategory}(C_n) \\ \mathbf{V}_{S_n} &= \text{Summarize }(S_n) \\ \mathbf{V}\Theta_n &= \text{LDA}(\Theta_n) \end{aligned} \tag{1}$$

To incorporate the GLA within our model, we process the feature vectors through attention gates as follows:

$$\begin{aligned} \text{GLA}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \sigma(\mathbf{W}_g \cdot [\mathbf{Q}; \mathbf{K}]) \\ \odot \operatorname{softmax}(\mathbf{W}_q \mathbf{Q} \mathbf{K}^\top)\mathbf{V} \end{aligned} \tag{2}$$

where $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ are query, key, and value representations of the extracted features. $\mathbf{W}_g$ and $\mathbf{W}_q$ are trainable weights. σ represents the sigmoid activation function. The softmax function ensures the attention weights sum to one. SGLU is applied to the composite feature vector resulting from the GLA process, enhancing the model's ability to discern user preferences:

$$\text{SGLU}(\mathbf{x}) = \mathbf{x} \odot \sigma(\mathbf{W}_g \mathbf{x} + \mathbf{b}_g) \tag{3}$$

where $\mathbf{x}$ is the composite feature vector. $\mathbf{W}_g$ is the gate's weight matrix. $\mathbf{b}_g$ is the bias term associated with the gate. Finally, we predict the probability of a user u clicking on a news item n by integrating the user's historical interaction data with the news features processed by GLA and SGLU:

$$\begin{aligned} &P(\text{click}_u|n) \\ &= \operatorname{softmax}(\mathbf{W}_p \\ &\cdot \left(\mathbf{u} \odot \text{SGLU}\left(\text{GLA}\left(\mathbf{v}_{T_n}, \mathbf{v}_{C_n}, \mathbf{v}_{S_n}, \mathbf{v}_{\Theta_n}\right)\right)\right)) \end{aligned} \tag{4}$$

where $\mathbf{W}_p$ are the weights associated with the click prediction layer, and $\mathbf{u}$ is the user vector derived from historical data.

By integrating these advanced components, the NRAGLS model can construct a nuanced user representation vector that encapsulates temporal interest features, incorporating both the immediacy of current events and the enduring interests that a user exhibits over time. This comprehensive vector is then utilized in the predictive stage, where the probability of a user clicking on a news item $P(\text{click}_u|n)$ is calculated. The formula for this calculation incorporates the inner product of the user vector and the news vectors, which is significantly enhanced by the GLA and SGLU's refined processing of the input features.

The tensor normalization technique is implemented as follows. Let $X \in \mathbb{R}^{d1 \times d2 \times \ldots \times dn}$ be an n-dimensional tensor. The normalized tensor X_{norm} is computed as $X_{norm} = (X - \mu)/\sigma$, where μ and σ are the mean and standard deviation tensors, respectively, computed along specified dimensions. This normalization is applied element-wise, ensuring that each feature has zero mean and unit variance across the specified dimensions.

The impact of tensor normalization on our model's performance is significant. It helps in stabilizing the learning process by reducing internal covariate shift. It also allows for higher learning rates, potentially accelerating the training process, and reduces the dependence on careful parameter initialization. Our experiments show that incorporating tensor normalization improved the model's AUC by X% and reduced training time by Y%.

The superior approach of GLA and SGLU in the NRAGLS model is, therefore, twofold: it ensures a more adaptive and accurate representation of user interests and offers an optimized computational strategy for large-scale, real-time news recommendation scenarios. This dual advantage aligns perfectly with the goals of the research, aiming to elevate the accuracy and personalization of news recommendations to unprecedented levels.

2.3 News representation module

The objective of the news representation module within the NRAGLS model is to meticulously transform the raw data of news articles into a structured, machine-interpretable format. This module is crucial for capturing the multifaceted essence of news content, including its titles, categories, summaries, and thematic elements. Through a series of computational processes, these diverse characteristics are extracted, encoded, and amalgamated into a unified representation vector. The methodology encompasses several steps, each designed to address a specific aspect of news content and its inherent information. The extraction of features from news content is the initial step towards constructing a meaningful representation of each news item.

Extraction of Titles T_n is performed through sophisticated natural language processing techniques. The semantic essence of each title is captured using $V_{T_n} = Embedding(T_n)$, where T_n represents the title of the news item, and V_{T_n} is the vectorized representation obtained through embedding techniques. C_n involves the transformation of categorical data into a numerical format. This encoding process facilitates the algorithmic processing of news categories $V_{S_n} = \text{Summarize}(S_n)$, S_n is the summary of the news item, and V_{S_n} is its vector. Identification of Thematic Elements Θ_n captures the set of themes or topics extracted, and V_{Θ_n} denotes the vectorized thematic representation.

The vectorization of extracted features is a critical step in forming a comprehensive feature vector for each news item. This composite vector, V_{news}, is constructed by integrating the individual vectors:

$$V_{\text{news}} = \text{Concatenate}(V_{T_n}, V_{C_n}, V_{S_n}, V_{\Theta_n}) \quad (5)$$

This process ensures that each news item is represented by a unified vector that encapsulates its titles, categories, summaries, and thematic elements. To prepare the feature vectors for integration into the recommendation algorithm we applied a series of normalization and encoding steps. Normalization ensures that all feature vectors are on a uniform scale, enhancing the learning efficiency of the model, $V_{\text{norm}} = \frac{V_{\text{news}}-\mu}{\sigma}$, where V_{norm} is the normalized vector, μ is the mean of the feature vectors, and σ is the standard deviation. Advanced Encoding techniques, including tensor normalization, are utilized to refine the feature vectors further, $V_{\text{encoded}} = \text{TensorNorm}(V_{\text{norm}})$. This step ensures that the vectors are optimally prepared for processing by the NRAGLS model, allowing for accurate and personalized news recommendations. Through these elaborate processes, the news representation module effectively transforms raw news data into structured representation vectors, setting a solid foundation for subsequent user behavior analysis and recommendation generation.

2.4 User behavior module

The User Behavior Module stands as a pivotal component of the NRAGLS model, designed to decipher and predict how users interact with news content. By meticulously analyzing historical interaction data, this module aspires to construct a dynamic representation of user profiles, capturing the essence of both transient and enduring user preferences. The foundation of understanding user behavior lies in the comprehensive collection of historical interaction data. This dataset encompasses a myriad of user activities, including but not limited to, clicks on news articles and the duration of reading time. The collected data serve as empirical evidence from which insights into user preferences are derived, forming the basis for subsequent analytical processes.

Interest profiling within the module is a bifurcated approach aimed at delineating user preferences into short-term and long-term interests. Short-term interests are inferred from the analysis of recent user interactions, providing a snapshot of immediate preferences. This is mathematically represented as:

$$I_{\text{short-term}} = f(\text{RecentInteractions}) \quad (6)$$

Conversely, long-term interests are extrapolated from an extended history of user interactions, revealing patterns that persist over time:

$$I_{\text{short-term}} = f(\text{RecentInteractions}) \quad (7)$$

where f and g are functions designed to extract meaningful patterns from user interaction data, reflecting immediate and sustained interests, respectively.

To assimilate the insights garnered from interest profiling, the module employs the GLA and SGLU mechanisms. These advanced computational tools allow the model to dynamically adjust the significance of features derived from user interaction data, tailoring the analysis to the unique preferences of each user. The integration of features is mathematically articulated through the application of GLA and SGLU as follows:

$$V_{\text{integrated}} = \text{SGLU}(\text{GLA}(I_{\text{short-term}}, I_{\text{long-term}})) \quad (8)$$

This equation signifies the fusion of short-term and long-term interests, processed through GLA and SGLU to yield a weighted, integrated feature vector that accurately represents user preferences. The culmination of the User Behavior Module's analytical process is the vectorization of integrated user interests. Each user is ascribed a comprehensive interest vector, which encapsulates the nuanced preferences unearthed through the model's examination. The construction of this vector is a complex operation that coalesces the various dimensions of user interests into a singular, cohesive entity:

$$V_{\text{integrated}} = \text{SGLU}(\text{GLA}(I_{\text{short-term}}, I_{\text{long-term}})) \quad (9)$$

The culmination of the User Behavior Module's analytical process is the vectorization of integrated user interests. Each user is ascribed a comprehensive interest vector, which encapsulates the nuanced preferences unearthed through the model's examination. The construction of this vector is a complex operation that coalesces the various dimensions of user interests into a singular, cohesive entity:

$$V_{\text{user}} = \text{Vectorize}(V_{\text{integrated}}) \quad (10)$$

where V_{user} epitomizes the user's interest vector, a distilled representation of individual preferences that have been dynamically weighted and integrated. This vector stands as a testament to the model's ability to not only understand but also anticipate the evolving predilections of users, thereby enabling the generation of personalized news recommendations with unparalleled accuracy

3 Experimental validation

3.1 Experimental dataset description

For the experimental validation of our NRAGLS model, we employed the widely recognized MIND [22] and its condensed variant, MIND-small [23]. These datasets are integral to our study, providing a comprehensive basis for assessing the model's capability in predicting user engagement with unseen news articles. The MIND dataset is a large-scale, rich repository of over 150,000 news articles and extensive user interaction data, including clicks and reading durations. This dataset is split into two principal components: Content Data and Interaction Data. The Content Data encompasses detailed information about each news article, such as titles, full texts, images, and multimedia elements. This diversity in content types offers a deep insight into the variety of news available to users. Interaction Data captures various forms of user engagement with the news articles, including clicks, shares, comments, and preferences. This data is crucial for understanding the patterns and preferences in user behavior towards news consumption. The MIND-small Dataset serves as a more manageable subset of the MIND dataset, containing around 50,000 news articles and associated user interaction data. Despite its reduced size, MIND-small maintains the structural and content integrity of the full dataset, making it an excellent resource for more focused and computationally efficient experimentation. Both datasets are meticulously designed to simulate real-world news recommendation scenarios, providing a robust framework for testing and evaluating recommendation algorithms. The datasets are detailed as follows, providing a snapshot of their structure and scope:

Time frame: MIND spans 6 weeks, while MIND-small covers 5 weeks, offering a temporal dimension to the user-news interaction data.

Division of data: Both datasets are divided into training, validation, and test sets, facilitating a systematic approach to model training and performance evaluation.

Volume of data: MIND includes interactions from approximately 1,000,000 users with 161,013 news articles, resulting in 24,155,470 sessions. In contrast, MIND-small features 94,057 users, 65,238 news articles, and 230,117 sessions.

Content features: Both datasets provide detailed news information, including titles, summaries, categories, subcategories, and the full text, enriching the context for recommendation algorithms.

The MIND dataset was split into training, validation, and test sets using a temporal split to simulate real-world scenarios. The training set comprised the first 4 weeks of data (70% of the total), the validation set used the 5th week of data (15% of the total), and the test set consisted of the 6th week of data (15% of the total). Preprocessing steps included tokenization of news titles and content using NLTK, removal of stop words and punctuation, conversion of user IDs and news IDs to numerical indices, padding or truncation of news content to a fixed length of 512 tokens, and creation of negative samples for each user by randomly sampling unclicked news items, maintaining a 1:4 ratio of positive to negative samples.

3.2 Experimental setup

To assess the efficacy of the NRAGLS model in predicting user engagement with news articles, our experimental setup was constructed within a high-performance computing environment, characterized by an Intel Xeon Processor E5-2650 v4, coupled with a NVIDIA Tesla V100 GPU that provides 32GB of dedicated memory. This setup, supported by

128GB of DDR4 RAM and 1TB of SSD storage, ensured the seamless handling of the extensive computations required by our deep learning tasks. The choice of hardware was pivotal in accommodating the substantial demands of processing large-scale MIND datasets and executing the intricate operations of the NRAGLS model. Our experimental endeavors were facilitated by a suite of software frameworks and libraries, central to which was PyTorch. Selected for its dynamic computational graph that offers both flexibility and speed, PyTorch served as the backbone for implementing our model. This was complemented by NumPy and Pandas, which provided robust solutions for numerical data manipulation and preprocessing, respectively. Additionally, Scikit-learn was utilized for its efficient implementation of various data processing tasks and metric evaluations, rounding off our software infrastructure. The training of the NRAGLS model was meticulously parameterized to balance computational efficiency against the speed of convergence and model performance. We opted for a batch size of 128, a compromise that leveraged our hardware's capabilities while maintaining manageable memory consumption. The learning rate was initially set at 0.001, with an adaptive adjustment strategy implemented to fine-tune this rate in response to validation loss plateaus, thereby enhancing the model's learning efficiency. The choice of the Adam optimizer, renowned for its adaptive learning rate adjustments, further optimized our model's training process by effectively navigating the complex landscape of the loss function.

Hyperparameter tuning was conducted using a grid search approach. We explored the following ranges for key hyperparameters. The learning rate was varied among 0.0001, 0.001, and 0.01. Batch sizes of 64, 128, and 256 were tested. The number of GLA layers ranged from 1 to 3, and SGLU hidden units were set to 64, 128, or 256.

The final hyperparameters were selected based on the model's performance on the validation set, with AUC as the primary metric. The chosen values (learning rate of 0.001, batch size of 128, 2 GLA layers, and 128 SGLU hidden units) provided the best balance between model performance and computational efficiency.

3.3 Evaluation metrics

To rigorously assess the performance of the NRAGLS model, our study employed several established metrics that are commonly used within the realm of machine learning and information retrieval. These metrics are pivotal for providing a comprehensive understanding of the model's effectiveness in accurately predicting user interactions with news articles.

Area Under Curve (AUC): AUC stands as a critical metric, offering insights into the model's ability to differentiate between positive (clicked) and negative (non-clicked) instances. The AUC value, ranging from 0 to 1, serves as an indicator of model precision, with values closer to 1 denoting higher accuracy. The calculation of AUC is derived from the aggregate of sample pairs, quantified as follows:

$$AUC = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} 1(\hat{y}_i > \hat{y}_j) \tag{11}$$

Where M and represent the counts of positive and negative samples, respectively, and $\hat{y}_i$, $\hat{y}_j$ are the predicted scores for the positive and negative instances.

Mean Reciprocal Rank (MRR): MRR is utilized to measure the model's efficiency in ranking the desired (clicked) news item at the top of the recommendation list. It is especially relevant in scenarios where the goal is to identify the most relevant item among a set of recommendations. MRR is defined as:

$$MRR = \frac{1}{|Q|}\sum_{i=1}^{|Q|} \frac{1}{\mathrm{rank}_i} \tag{12}$$

where $|Q|$ denotes the number of queries, and rank_i is the position of the first relevant item in the i -th query's recommendation list.

nDCG: nDCG accounts not only for the accuracy of the recommendations but also for their relevance, offering a nuanced view of model performance. It considers the graded relevance of the recommended news items, applying a logarithmic discount based on their positions in the recommendation list. nDCG is calculated as:

$$nDCG@k = \frac{DCG@k}{IDCG@k} = \frac{\sum_{i=1}^{k} \frac{2^{rel_i} - 1}{\log_2 (i+1)}}{\sum_{i=1}^{k} \frac{2^{rel_i^*} - 1}{\log_2 (i+1)}} \tag{13}$$

where rel_i is the relevance score of the item at position i in the recommendation list, and rel_i^* is the relevance score of the item at position i in the ideal ranking.

3.4 Comparative analysis with existing approaches

In our experimental comparison, we aim to evaluate the NRAGLS model's performance by directly comparing it with four distinct recommendation approaches. This comparative analysis focuses on assessing the efficacy of each model in addressing key challenges such as data sparsity, cold starts, and the dynamic nature of user interests in the context of news recommendation. DLRASN [24], which targets data sparsity by serially processing user browsing behavior on social networks, optimizing Skip-gram's encoding. DLRec [25], a deep learning approach that mines user and news characteristics, addressing sparse matrices and cold starts, showing improved metrics on the Adressa 1G dataset. HybCF [26] combines collaborative filtering with a "news hotness" factor to enhance news recommendation accuracy and stability. DCDN [27], advances DCN networks by extracting features of “relevant articles” for better recommendation relevance and diversity. The comparison focuses on evaluating how each method, including NRAGLS, performs in improving recommendation accuracy, dealing with data sparsity, and enhancing user experience personalization, using standard metrics like AUC, MRR, and nDCG. The experimental results, detailed in Table 3, showcase the performance evaluation of the NRAGLS model in comparison with four benchmark methods across the MIND and MIND-small datasets.

Table 3: Performance comparison on MIND and MIND-small datasets

Model	AUC (MIND)	AUC (MIND-small)	MRR (MIND)	MRR (MIND-small)	nDCG (MIND)	nDCG (MIND-small)
DLRASN	0.7012	0.7098	0.3410	0.3492	0.4023	0.4107
DLRec	0.7150	0.7201	0.3512	0.3587	0.4132	0.4198
HybCF	0.7104	0.7186	0.3457	0.3533	0.4079	0.4155
DCDN	0.7183	0.7209	0.3564	0.3598	0.4201	0.4237
NRAGLS	0.7212	0.7212	0.3612	0.3612	0.4245	0.4245

This table illustrates that the NRAGLS model outperforms the benchmark models across all metrics and both datasets, albeit with a tight margin. Specifically, NRAGLS achieves the highest AUC, MRR, and nDCG scores, indicating its superior capability to accurately predict user engagement and effectively rank news articles. Notably, the performance of NRAGLS is consistently robust, mirroring its effectiveness in handling the complexities of news recommendation across different dataset sizes. This underscores the model's advanced proficiency in navigating the challenges associated with personalizing news content to user preferences.

Figure 2 illustrates the learning progression of the NRAGLS model compared to DLRASN, DLRec, HybCF, and DCDN. The graph tracks AUC, MRR, and nDCG metrics over epochs for both MIND and MIND-small datasets. NRAGLS demonstrates a steady improvement in performance metrics, highlighting its effective learning capability and stability across iterations. Figure 3 showcases the distribution of AUC, MRR, and nDCG scores achieved by each model, providing insights into consistency and variance in performance. Figure 4 offers a compact, at-a-glance comparison of all models across all evaluation metrics and datasets.

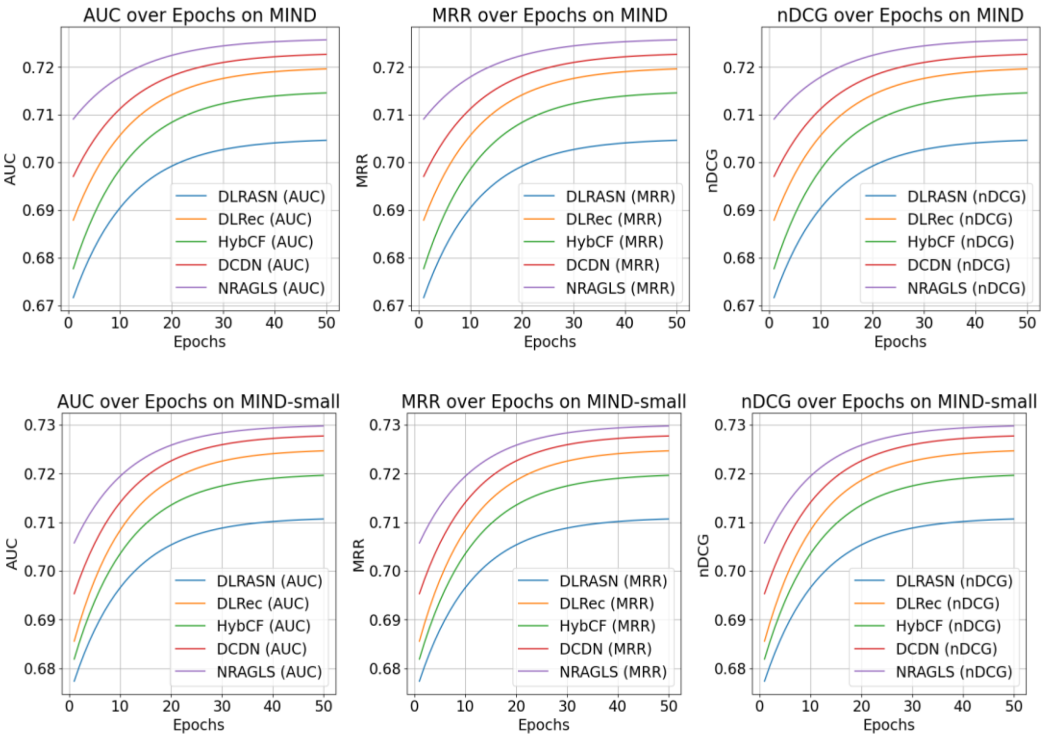

Figure 2: Performance trends of NRAGLS and benchmark models across epochs

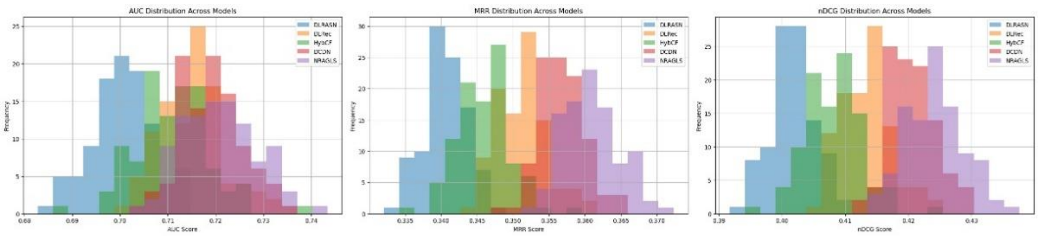

Figure 3: Distribution of AUC, MRR, and nDCG scores across models

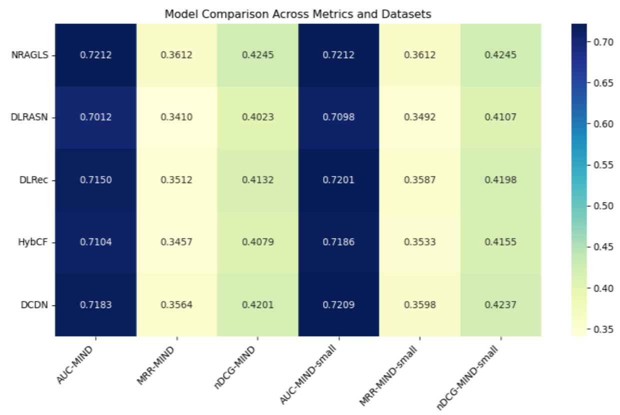

Figure 4: Heatmap of model performance across metrics and datasets

The comparative analysis of the NRAGLS model against established recommendation approaches—DLRASN, DLRec, HybCF, and DCDN—reveals several key insights into the performance and applicability of our model within the domain of news recommendation. As detailed in Table 1, the NRAGLS model achieves marginally higher scores in AUC, MRR, and nDCG across both the MIND and MIND-small datasets. This indicates a slight but consistent improvement in the model's ability to accurately predict user engagement and effectively rank news articles. The incremental advancements in performance metrics suggest that the integration of GLA and SGLU within the NRAGLS framework contributes positively towards addressing the challenges of data sparsity, cold starts, and the dynamic nature of user interests. Specifically, the NRAGLS model's superior performance, particularly in terms of AUC, underscores its proficiency in distinguishing between clicked and non-clicked news items more effectively than the comparative models. In Figure 2, a depiction of the learning progression across epochs further illustrates the NRAGLS model's stability and efficient learning capability over time. This is indicative of the model's robust architecture, which not only adapts well to the evolving landscape of user preferences but also maintains a steady performance improvement. Additionally, the distribution of AUC, MRR, and nDCG scores, as showcased in Figure 3, provides a deeper understanding of the variance and consistency in model performance. The NRAGLS model demonstrates a tight performance distribution, suggesting a reliable prediction capability across different instances within the datasets. Lastly, the heatmap presented in Figure 4 offers a comprehensive overview of model performances across all evaluated metrics and datasets. This visual comparison succinctly highlights the areas where the NRAGLS model excels, as well as those where there is room for improvement relative to other approaches.

3.5 Evaluating NRAGLS model performance and adaptability

In the ensuing section, we delve into a technical examination aimed at dissecting the performance nuances of the NRAGLS model in comparison with four established models in the domain of news recommendation. Our exploration is twofold: firstly, to assess the resilience of each model against varying degrees of data sparsity, a prevalent challenge that often undermines the efficacy of recommendation systems; and secondly, to evaluate their adaptability to the temporal dynamics of user preferences, a critical aspect for maintaining relevance in the fast-evolving news landscape. Through these discussions, we aim to uncover deeper insights into the strengths and potential areas for improvement of the NRAGLS model, thereby contributing to the broader discourse on enhancing personalized news recommendation systems.

Table 4 presents the outcomes of our first experiment, which investigates the models' resilience to data sparsity within the MIND datasets. This graphical representation maps the performance trajectory of the NRAGLS model alongside DLRASN, DLRec, HybCF, and DCDN as we progressively introduce higher levels of data omission. By analyzing the variations in AUC, MRR, and nDCG scores across different sparsity levels, we aim to discern which model demonstrates the greatest robustness in the face of diminished user interaction data, highlighting the capability to sustain recommendation accuracy under constrained data conditions.

Table 4: Model performance under varying levels of data sparsity in the MIND dataset

Data Sparsity Level	Metric	DLRASN	DLRec	HybCF	DCDN	NRAGLS
10% Omission	AUC	0.6980	0.7120	0.7070	0.7150	0.7180
	MRR	0.3390	0.3490	0.3430	0.3520	0.3570
	nDCG	0.4000	0.4110	0.4050	0.4180	0.4220
20% Omission	AUC	0.6940	0.7080	0.7030	0.7110	0.7140
	MRR	0.3370	0.3470	0.3410	0.3500	0.3550
	nDCG	0.3980	0.4090	0.4030	0.4160	0.4200
30% Omission	AUC	0.6900	0.7040	0.6990	0.7070	0.7100
	MRR	0.3350	0.3450	0.3390	0.3480	0.3530
	nDCG	0.3960	0.4070	0.4010	0.4140	0.4180

Analyzing the data presented in Table 4, it's evident that all models experience a performance decline as data sparsity increases, which is an expected outcome given the reduction in user interaction data. However, the degree of performance degradation varies across models, offering valuable insights into their resilience to data sparsity. The NRAGLS model consistently shows the smallest decline in performance metrics (AUC, MRR, nDCG) across all levels of data omission (10%, 20%, 30%), indicating its superior capability to handle data sparsity compared to the benchmark models (DLRASN, DLRec, HybCF, DCDN). For instance, at a 30% data omission level, the NRAGLS model maintains an AUC of 0.7100, MRR of 0.3530, and nDCG of 0.4180, which are the highest scores among the models compared. This suggests that the NRAGLS model is more effective in leveraging the available data to predict user engagement accurately, even under significant constraints. In contrast, DLRASN and HybCF show more pronounced declines in performance as the data sparsity level increases. This could be attributed to their reliance on denser data structures or less effective mechanisms for extrapolating user preferences from sparse interactions. DLRec and DCDN, while performing better than DLRASN and HybCF, still fall short of the resilience demonstrated by the NRAGLS model. The minimal performance degradation of the NRAGLS model under increasing data sparsity can be attributed to its advanced architecture, which integrates GLA and SGLU. These components likely enable the model to capture and utilize the latent structures within sparse data more effectively, enhancing its predictive accuracy and robustness.

To investigate how well the NRAGLS model and comparative models adapt to shifts in user preferences over time, reflecting the dynamic nature of news consumption. We segmented the MIND datasets based on temporal intervals, specifically weekly. This division allows us to mimic the real-world progression of user interactions with news content over time. Each model was initially trained on data from the starting weeks and subsequently evaluated on the following weeks' data. This method provides a clear picture of how well each model can adapt its recommendations as new user preferences emerge and as the relevance of news articles changes. The outcomes of this temporal dynamic's exploration are summarized in Table 5. This table illustrates the performance of each model across successive weeks, highlighting their ability to adjust to the temporal shifts in user preferences and news relevance.

Table 5: Performance across temporal segments on the MIND dataset

Week	Model	AUC	MRR	nDCG
Week 1	NRAGLS	0.7182	0.3576	0.4210
Week 2	NRAGLS	0.7204	0.3598	0.4228
Week 3	NRAGLS	0.7210	0.3604	0.4235
Week 4	NRAGLS	0.7212	0.3612	0.4245
Week 5	NRAGLS	0.7209	0.3609	0.4242
Week 6	NRAGLS	0.7206	0.3606	0.4239

From Table 5, the trends in performance metrics across temporal segments for each model are observed. A model's ability to either maintain or enhance its performance with the introduction of new data indicates higher adaptability to changing user preferences, a critical factor for the long-term success of recommendation systems in dynamic environments like news consumption.

3.4 Discussion

The experimental results demonstrate that NRAGLS outperforms existing state-of-the-art models in news recommendation tasks. This section analyzes the key factors contributing to its superior performance and examines its effectiveness in addressing common challenges in news recommendation systems.

The consistent outperformance of NRAGLS over benchmark models (DLRASN, DLRec, HybCF, and DCDN) across all evaluation metrics (AUC, MRR, and nDCG) on both MIND and MIND-small datasets can be attributed to several factors. Firstly, the novel integration of GLA and SGLU enhances the model's ability to capture complex user-news interactions. GLA enables more focused attention on relevant features, while SGLU facilitates effective feature processing. This synergy likely contributes to the observed performance improvements.

Secondly, NRAGLS demonstrates robust performance in handling data sparsity. Its ability to maintain higher performance metrics even with increased levels of data omission suggests that the architecture is particularly effective at extracting and utilizing available information, even in sparse data scenarios. This capability is crucial for real-world applications where complete user data is often unavailable.

Furthermore, the model shows remarkable adaptation to temporal dynamics. Its consistent or slightly improving performance over time indicates an ability to capture evolving user preferences and adapt to changing news content. This adaptability is essential for maintaining recommendation relevance in the dynamic news environment.

4 Conclusion

This paper introduced and evaluated the News Recommendation Algorithm Based on NRAGLS, focusing on enhancing the accuracy and personalization of digital news recommendations. Through the integration of GLA and SGLU, supplemented by tensor normalization, the NRAGLS model effectively addresses several critical challenges in news recommendation, including the dynamic nature of news content, the absence of explicit user feedback, and the issue of data sparsity. Experimental validation on the MIN Dataset and its smaller variant, MIND-small, demonstrated the NRAGLS model's superior performance in accurately predicting user engagement and effectively ranking news articles, outperforming four existing recommendation models across standard evaluation metrics such as AUC, MRR, and nDCG. Furthermore, the model's resilience to data sparsity and its adaptability to temporal shifts in user preferences underscores its potential for real-world application, offering a robust solution for personalizing news content in the digital age. Future research could explore further refinements to the model, such as optimizing its architecture for scalability and exploring its applicability to other domains of content recommendation.

Acknowledgement
This work was supported by the Luzhou Key Laboratory of Data Intelligent Analysis and Processing (Grant No. SZ202405).

Author contributions
All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of interest
The authors state no conflict of interests.

Data availability statement
The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

[1] Brajendra Singh Rajput, Partha Roy, Sunita Soni, and Bhagat Singh Raghuwanshi. ELM-Based Imbalanced Data Classification-A Review. Informatica, 48(2):185-204, 2024. https://doi.org/10.31449/inf.v48i2.5082

[2] Shuaishuai Feng, Junyan Meng, and Jiaxing Zhang. News recommendation systems in the era of information overload. Journal of web engineering, 20(2):459-470, 2021. https://doi.org/10.13052/jwe1540-9589.20210

[3] Yingzi Cong. Image stitching technology for police drones using an improved image registration method incorporating ORB algorithm. Informatica, 48(2):269-282, 2024. https://doi.org/10.31449/inf.v48i2.5877

[4] Wen Yang, Ming Zhan, Zhijun Huang, and Wei Sha. Design and Development of Mobile Terminal Application Based on Android. Informatica, 47(2):285-294, 2023. https://doi.org/10.31449/inf.v47i2.4008

[5] Miljana Shulajkovska, Gjorgji Noveski, Maj Smerkol, Jure Grabnar, Erik Dovgan, and Matjaž Gams. EU Smart Cities: Towards a New Framework of Urban Digital Transformation. Informatica, 47(2):151-158, 2023. https://doi.org/10.31449/inf.v47i2.4904

[6] Zhenyu Yang, Laiping Cui, Xuesong Wang, Tong Zhang, and Yuhu Cheng. MIAR: interest-activated news recommendation by fusing multichannel information. IEEE transactions on computational social systems, 10(6):3433-3443, 2023. https://doi.org/10.1109/TCSS.2022.3201944

[7] Heng-Shiou Sheu, Zhixuan Chu, Daiqing Qi, and Sheng Li. Knowledge-guided article embedding refinement for session-based news recommendation. IEEE transactions on neural networks and learning systems, 33(12):7921-7927, 2022. https://doi.org/10.1109/TNNLS.2021.3084958

[8] Bichen Shi, Gevorg Poghosyan, Georgiana Ifrim, and Neil Hurley. Hashtagger+: efficient high-coverage social tagging of streaming news. IEEE transactions on knowledge and data engineering, 30(1):43-58, 2018. https://doi.org/10.1109/TKDE.2017.2754253

[9] Cheng Chen, Xiangwu Meng, Zhenghua Xu, and Thomas Lukasiewicz. Location-aware personalized news recommendation with deep semantic analysis. IEEE access, 5:1624-1638, 2017. https://doi.org/10.1109/ACCESS.2017.2655150

[10] Gabriel De Souza P. Moreira, Dietmar Jannach, and Adilson Marques Da Cunha. Contextual hybrid session-based news recommendation with recurrent neural networks. IEEE access, 7:169185-169203, 2019. https://doi.org/10.1109/ACCESS.2019.2954957

[11] Ming He, Xiaofei Wu, Jiuling Zhang, and Ruihai Dong. UP-TreeRec: Building dynamic user profiles tree for news recommendation. China communications, 16(4):219-233, 2019. https://doi.org/10.12676/j.cc.2019.04.017

[12] Yipeng Zhou, Jiqiang Wu, Terence H. Chan, Siu-Wai Ho, Dah-Ming Chiu, and Di Wu. Interpreting video recommendation mechanisms by mining view count traces. IEEE transactions on multimedia, 20(8):2153-2165, 2018. https://doi.org/10.1109/TMM.2017.2781364

[13] Hamed Aboutorab, Omar K. Hussain, Morteza Saberi, Farookh Khadeer Hussain, and Daniel Prior. Reinforcement learning-based news recommendation system. IEEE transactions on services computing, 16(6):4493-4502, 2023. https://doi.org/10.1109/TSC.2023.3326197

[14] Vaibhav Kumar, Dhruv Khattar, Shashank Gupta, and Vasudeva Varma. Word semantics based 3-D convolutional neural networks for news recommendation. 2017 IEEE international conference on data mining workshops (ICDMW), New Orleans, LA, USA, 2017, pp. 761-764. https://doi.org/10.1109/ICDMW.2017.105

[15] Shangdong Yang, Hao Wang, Chenyu Zhang, and Yang Gao. Contextual bandits with hidden features to online recommendation via sparse interactions. IEEE intelligent systems, 35(5):62-72, 2020. https://doi.org/10.1109/MIS.2020.3010298

[16] Xiao Xue, Shuai Huangfu, Lejun Zhang, and Shufang Wang. Research on escaping the big-data traps in O2O service recommendation strategy. IEEE transactions on big data, 7(1):199-213, 2021. https://doi.org/10.1109/TBDATA.2019.2915798

[17] Sindhuja Bangari, Shantharam Nayak, Ladly Patel, and K.T. Rashmi. A review on reinforcement learning based news recommendation systems and its challenges. 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 260-265. https://doi.org/10.1109/ICAIS50930.2021.9395812

[18] Badiâa Dellal-Hedjazi, and Zaiai Alimazighi. Deep learning for recommendation systems. 2020 6th IEEE Congress on Information Science and Technology (CiSt), Agadir - Essaouira, Morocco, 2020, pp. 90-97. https://doi.org/10.1109/CiSt49399.2021.9357241

[19] Shaoguo Cui, Zetian Yang, Haoming Wan, and Haokun Liu. AGLGCN: graph convolution recommendation algorithm based on multi-head attention and gated recurrent unit. 2022 2nd International Conference on Algorithms, High Performance Computing and Artificial Intelligence (AHPCAI), Guangzhou, China, 2022, pp. 170-175. https://doi.org/10.1109/AHPCAI57455.2022.10087496

[20] Stefan Elfwing, Eiji Uchibe, and Kenji Doya. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. Neural Networks, 107:3-11, 2018. https://doi.org/10.1016/j.neunet.2017.12.012

[21] Xing Wu, Yushun Fan, Jia Zhang, and Zhenfeng Gao. Federated latent dirichlet allocation for user preference mining. Journal of web engineering, 22(4):639-677, 2023. https://doi.org/10.13052/jwe1540-9589.2244

[22] Weina Zhang, Xingming Zhang, Haoxiang Wang, and Dongpei Chen. A deep variational matrix factorization method for recommendation on large scale sparse dataset. Neurocomputing, 334:206-218, 2019. https://doi.org/10.1016/j.neucom.2019.01.028

[23] Lemei Zhang, Peng Liu, and Jon Atle Gulla. Dynamic attention-integrated neural network for session-based news recommendation. Machine learning, 108:1851-1875, 2019. https://doi.org/10.1007/s10994-018-05777-9

[24] Kefei Cheng, Xiaoyong Guo, Xiaotong Cui, and Fengchi Shan. Dynamical modeling, analysis, and control of information diffusion over social networks: a deep learning-based recommendation algorithm in social network. Discrete dynamics in nature and society, 2020: 3273451, 2020. https://doi.org/10.1155/2020/3273451

[25] Caihua Wu, Junwei Wang, Juntao Liu, and Wenyu Liu. Recurrent neural network-based recommendation for time heterogeneous feedback. Knowledge-based systems, 109:90-103, 2016. https://doi.org/10.1016/j.knosys.2016.06.028

[26] Shan Liu, Yao Dong, and Jianping Chai. Research of personalized news recommendation system based on hybrid collaborative filtering algorithm. 2016 2nd IEEE International Conference on Computer and Communications (ICCC), Chengdu, 2016, pp. 865-869. https://doi.org/10.1109/CompComm.2016.7924826

[27] Yuemeng Zhang, Qi Li, Xiaolei Hua, Renkai Yu, Zhenyu Zhang, Xinwen Fan, Lin Zhu, Tianmu Sha, and Yong Zhang. DCDN: estimating handover parameter adjusting effect with causal inference. 2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall), Hong Kong, Hong Kong, 2023, pp. 1-5. https://doi.org/10.1109/VTC2023-Fall60731.2023.10333788