GLoCIM: Global-view Long Chain Interest Modeling for news recommendation

Zhen Yang[1], Wenhui Wang[1*], Tao Qi[2], Peng Zhang[1],
Tianyun Zhang[1], Ru Zhang[1], Jianyi Liu[1], Yongfeng Huang[2,3]

[1]School of Cyberspace Security, Beijing University of Posts and Telecommunications
[2]Department of Electronic Engineering, Tsinghua University
[3]Zhongguancun Laboratory, Beijing, China
{yangzhenyz, edwin_wang, zhangpengzp, zhangtianyun, zhangru, liujy}@bupt.edu.cn
taoqi.qt@gmail.com   yfhuang@tsinghua.edu.cn

Abstract

Accurately recommending candidate news articles to users has always been the core challenge of news recommendation system. News recommendations often require modeling of user interest to match candidate news. Recent efforts have primarily focused on extracting local subgraph information in a global click graph constructed by the clicked news sequence of all users. However, the computational complexity of extracting global click graph information has hindered the ability to utilize far-reaching linkage which is hidden between two distant nodes in global click graph collaboratively among similar users. To overcome the problem above, we propose a **G**lobal-view **Lo**ng **C**hain **I**nterests **M**odeling for news recommendation (**GLoCIM**), which combines neighbor interest with long chain interest distilled from a global click graph, leveraging the collaboration among similar users to enhance news recommendation. We therefore design a long chain selection algorithm and long chain interest encoder to obtain global-view long chain interest from the global click graph. We design a gated network to integrate long chain interest with neighbor interest to achieve the collaborative interest among similar users. Subsequently we aggregate it with local news category-enhanced representation to generate final user representation. Then candidate news representation can be formed to match user representation to achieve news recommendation. Experimental results on real-world datasets validate the effectiveness of our method to improve the performance of news recommendation.

1 Introduction

News recommendation, vital for aligning content with users' reading preferences (Das et al., 2007), requires accurate modeling of user interests given the daily production of news (Wang et al., 2018).
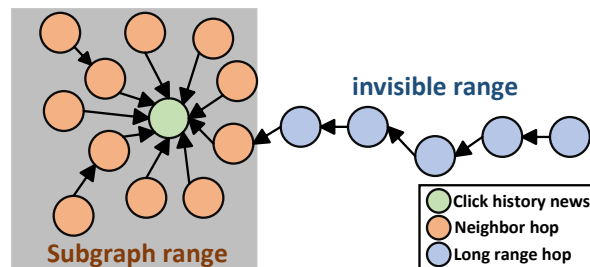
---

*Corresponding author.

Figure 1: An example of some news click global graph with traditional limited hop modeling without long chain selection and utilization.

In news recommendation, local modeling uses deep learning techniques like attention mechanisms, Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM) to extract semantic information from clicked news, generating user interest representations from local user click sequences for recommendations (Wu et al., 2019c).

Methods based on local user interest have quickly reached a bottleneck in news recommendation, prompting the adoption of GNNs. These enhance recommendations by using external knowledge graphs to capture implicit interests. For instance, KIM (Qi et al., 2021a) employs knowledge graphs to enrich representations of user history click news and candidate news.

However, these methods inadequately capture common interest features for collaborative recommendations among similar users. Recent GNN research focuses on building local subgraphs from a global click graph based on all users' click sequences, as shown in Figure 1. GLORY (Yang et al., 2023) builds a global click graph from all users' click sequences to provide a global-view enhancing recommendations. Commonly, local graphs are constructed for individual users and analyzed with various GNNs (Velickovic et al., 2017; Chen et al., 2020b; Li et al., 2016), but this approach incurs significant computational costs when

modeling far-reaching linkage between two distant nodes in global click graph.

It is well understood that two key barriers of computational complexity problem obstruct the full exploitation of the global click graph. Firstly, the process of distilling information from the global graph, based on all users' click history, often yields a graph that is excessively large, typically incurring significant efficiency costs in terms of memory and computational speed. Secondly, GNNs (Velickovic et al., 2017; Chen et al., 2020b; Li et al., 2016), which utilize subgraphs restrict short-range hop neighbors, exhibit limited expressive capabilities and face difficulties in capturing long chain interest. This limitation makes it challenging to effectively distill long chain interest from global click graph.

Inspired by metapath (Fan et al., 2019) from heterogeneous graph, we attempt to capture far-reaching linkage in the global click graph, thereby constructing the most valuable long chain for a user's each clicked news. We propose long chain selection algorithm and interest encoder to distill long chain interest from the global click graph. A gated network integrates these with neighbor interest for a collaborative interest representation, which is merged with local category enhanced news interest to form the user representation. A candidate neighbor entity encoder then forms a candidate representation. Our main contributions are:

- We design a long chain selection algorithm based on popularity, category and semantics to select long chains for each clicked news of all users in global click graph, so that we can select the most valuable long chains to better capture far-reaching linkage.

- We propose long chain attention network to distill the long chain interest at the news level and user level respectively, subsequently design a gated network to combine long chain interest and neighbor interest to enhance collaborative interest among users.

- Extensive experiments on real world datasets demonstrate that GLoCIM improves news recommendation effectiveness, outperforming other baseline methods.

2 Related Work

A common approach is to study local news semantic representation through user's reading history with some deep learning techniques. Several approaches (Okura et al., 2017; An et al., 2019a) have been refined to better model user interest based on their news interaction histories. LSTUR (An et al., 2019b) combines long-term interest represented by a user's ID with short interest derived from recent activities, while DAN (Zhu et al., 2019) enhances user hidden sequential features of user's clicks through attention-based parallel CNN. When the effectiveness of local semantic representation reaches its limitations, improving personalized recommendation can be accomplished by creating an extensive news graph that leverages the multi-dimensional data within it, User-as-graph (Wu et al., 2021) and DIGAT (Mao et al., 2022) employ graph-based approaches to capture higher-dimensional information by leveraging user behavior patterns and dual graph interactions. KIM (Qi et al., 2021a) uses a co-encoding method with a knowledge graph to understand the interactions between clicked and candidate news, enhancing the personalization of recommendations. HieRec (Qi et al., 2021b) subsequently models user interest from subtopics to a comprehensive synthesis within a hierarchical interest framework. This method reflects a trend towards a more layered understanding of user interest in news recommendation.

Drawing on techniques that employ user-news graphs for news recommendation, it has been demonstrated that graph modeling significantly enhances the effectiveness of recommendation systems (Chen et al., 2020a). GERL (Ge et al., 2020) and GUND (Hu et al., 2020b) formulate news and users jointly in a bipartite graph to model news-user interaction. Some methods also incorporate a global click graph derived from user click history to enrich collaborative information. GNewsRec (Hu et al., 2020a) introduces a heterogeneous graph model combined with GNNs and an attention-based LSTM to improve news recommendation. By integrating user, news, and topic interactions, the model distills both long-term and short-term user interest. GLORY (Yang et al., 2023) introduces a model that combines global and local insights for personalized recommendation. It uses a global-aware encoder to dive into a global click graph and gated GNN to enhance news representation via a historical aggregator. Nonetheless, these approaches have limitations in fully exploiting far-reaching linkage and optimizing the utilization of long chain interest among collaborative users, thus

unintentionally or inadequately overlooking the long chain interest distillation.

Method	(C1)	(C2)	(C3)	(C4)
LSTUR (An et al., 2019a), CUPMAR (Tran et al., 2021), DAN (Zhu et al., 2019), NRMS (Wu et al., 2019c) , Hi-eRec (Qi et al., 2021b)	✓	-	-	-
KIM (Qi et al., 2021a), DI-GAT (Mao et al., 2022), User-as-Graph (Wu et al., 2021)	✓	✓	-	-
GLORY (Yang et al., 2023), GNewsRec (Hu et al., 2020a)	✓	✓	✓	-
GLoCIM	✓	✓	✓	✓

Table 1: Comparison of methods based on various analytical dimensions: local modeling (C1), local graph assistance (C2), global graph utilization (C3), and global long chain interest distillation (C4).

3 Method

GLoCIM is mainly divided into four phases: global click graph construction and news semantic representation generation, long chain and neighbor subgraph selection, interest distillation, and user-news matching, which is shown in Figure 2. We mainly propose long chain selection algorithm to dynamically obtain long chains, and then design long chain interest encoder to distill long chain interest and gated neural network to fuse long chain interest with neighbor interest in global click graph.

Problem Formulation. The click history sequence of a user u can be denoted as $\mathbf{H}_u = [d_1, d_2, ..., d_H]$, where H is the number of historical news articles. Each news article d_i has a title, which contains a text sequence $\mathbf{T}_i = [w_1, w_2, ..., w_T]$ consisting of T word tokens, and an entity sequence, which is denoted by $\mathbf{E}_i = [e_1, e_2, ..., e_E]$ consisting of E entities. The objective is to predict the level of interest $s_{u,c}$ for a given candidate news article d_c and user u, which reflects the likelihood of a clicking action occurring between them.

3.1 Category-enhanced news Representation

In the first phase of news semantic representation generation, to model user behaviors to filter the most valuable long chains in global click graph, we leverage category information to enhance news representation. We limit our analysis to the last L_{his} clicked news in user u's history $\mathbf{H_u}$ and the candidate news d_c. Only the first L_{title} words of each news title are considered for input.

News Encoder. We initialize each news n with a word embedding layer using GloVe (Pennington et al., 2014) embeddings to convert news text into embedding vectors $\mathbf{x}_n$. These are processed by a multi-head self-attention (MSA) layer to produce each news semantic representations $\mathbf{X}_n$:

$$\mathbf{X}_n = \text{MSA}(\mathbf{x}_n) \tag{1}$$

Thus, the title representation $\mathbf{X}_n$ is enhanced with news category c_n and subcategory sc_n using concatenation and trainable W_{mix} and b_{mix} to generate news representation $\mathbf{C}_n$. We can integrate both sementics and category to make preparations for better matching between clicked news:

$$\mathbf{C}_n = (\mathbf{X}_n \oplus c_n \oplus sc_n) \cdot W_{\text{mix}} + b_{\text{mix}} \tag{2}$$

Local News encoder. In the third phase of semantic interest distillation, a text attention layer aggregates news final representations to form user local interest news representation $\mathbf{h}^{ln}$ as follows:

$$\mathbf{h}^{ln} = \sum_{i=1}^{L_{his}} \alpha_i \mathbf{C}_i \tag{3}$$

$$\alpha_i = \frac{exp(\mathbf{q}^\top \tanh(\mathbf{W} \cdot \mathbf{C}_n^i))}{\sum_{j=1}^{L_{his}} exp(\mathbf{q}^\top \tanh(\mathbf{W} \cdot \mathbf{C}_n^j))} \tag{4}$$

where $\mathbf{q}$ is a query vector, $\mathbf{W}$ is a trainable matrix.

We feed pre-trained TransE embeddings of entities in news into an entity self-attention network to form entity representation $\mathbf{X}_e$, and an attention network aggregates them into user local interest entity representation $\mathbf{h}^{le}$.

3.2 Long chain Interest modeling

To model long chain interest, we first select the most valuable long chains through long chain selection algorithm, and then use the long chain interest encoder to distill interest. After every certain number of local news encoder training steps in the above process, we use updated news representations to select updated news long chains and distill long chain interest, thereby modeling the final long chain interest. The details are shown below:

3.2.1 Long chain Selection algorithm

The long chain selection algorithm consists of alternative long chain selection and pruning from global click graph G as shown in Algorithm 1. G represents the relationships between news articles, constructed from the temporal sequence of click
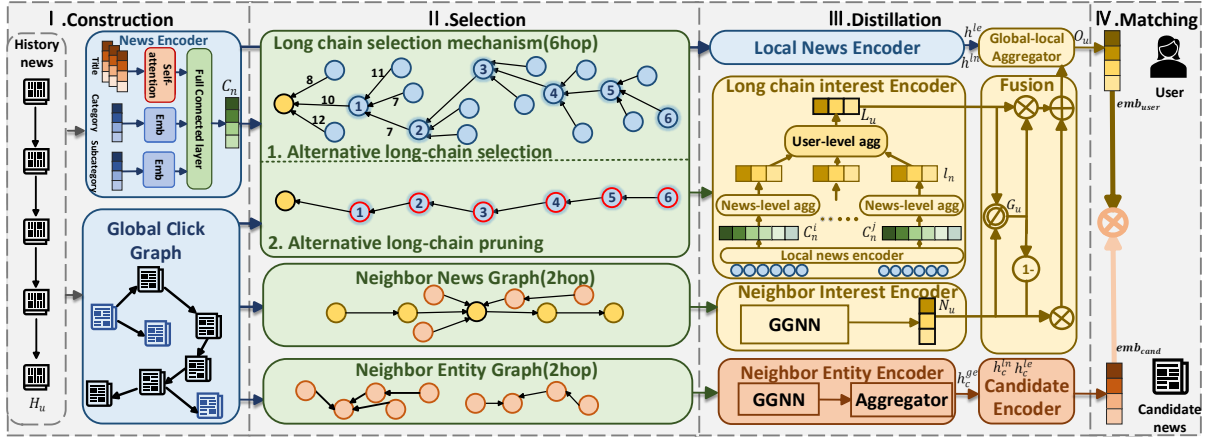
Figure 2: The overall architecture of our model. The graph construction is based on the user click history.

histories of all users. Since each user possesses its own click histories, G is subsequently constructed to facilitate recommendations.

Alternative long chain Selection. For a user u within the range from u_1 to u_N, a sequence $(n_A, n_B, n_C, \ldots)$ is formed in click timestamp sequence. For any two adjacent news article in the sequence, a directed pair (n_A, n_B) can be represented within a global click graph. For each directed pair (n_A, n_B), when the click sequence of user u $(n_{1,u}, n_{2,u}, ..., n_{i,u}, n_{i+1,u} \ldots)$ satisfies $n_{i,u} = n_A, n_{i+1,u} = n_B$, the click frequency of user u can be defined as $f_u(n_A \to n_B) = 1$; otherwise, $f_u(n_A \to n_B) = 0$.

In addition with semantics and category, our third long chain selecting strategy utilizes the temporal popularity of news clicks, making news directed pairs with more clicks more likely to be captured. Therefore, to accurately utilize the collaborative relationships among different users to match user interest, for the global click graph, different users may have a non-zero click frequency for the same directed pair. Thus, an accumulated frequency for a directed pair can be defined:

$$af(n_A \to n_B) = \sum_{i=u_1}^{u_N} f_i(n_A \to n_B) \quad (5)$$

We apply a similarity-based method to dynamically find the most relevant neighbors for each news article in G. For a target news n_t, $\forall n \in G, af(n \to n_t) \neq 0$, n is a neighbor of n_t.

Define a set $N_{n_t}^1 = \{n \in G \mid af(n \to n_t) \neq 0\}$ as the first hop neighbor set of n_t. Subsequently, we define $N_{n_t}^{i+1} = \{n_{t,i} \in N_{n_t}^i, n \in G \mid af(n \to n_{t,i}) \neq 0\}$ as the i+1 hop neighbor set of n_t. However, as the hop count i increases, the size of $N_{n_t}^{i+1}$

grows exponentially. Therefore, we select an alternative long chain (ALC_i) containing a small number of important neighbors from each hop's neighbor set. We select the alternative long chain with branch m using the top m click weight news in $N_{n_t}^1$. Here, the click weight is defined below:

$$W_j^1 = \frac{af(n_j \to n_t)}{\sum_{k=1}^{|N_{n_t}^1|} af(n_k \to n_t)}, n_j \in N_{n_t}^1 \quad (6)$$

Algorithm 1 Long Chain Selection

Input: Global click graph G, Target news n_t, Long chain length len, Alternative long chain branches m, Total training step st, The weight set of neighbor news at i hop is W^i, Total update times is k.

Output: Long chain for n_t:$LC(n_1, n_2, \ldots, n_{len})$

1: $LC \leftarrow \text{Queue}()$
2: **while** $\left(\text{step mod } \frac{st}{k}\right) = 0$ **do**
3:    $\forall n \in G$, Local News Encoder$(n_t) \to C_{n_t}$
4:    ALC_i's hop is $i, i \leftarrow 0$
5:    $n_c \leftarrow n_t$
6:    **while** $i < len$ **do**
7:      $i + +$
8:      $W^i \leftarrow caculate(N_{n_c}^1)$
9:      $\text{sort}(W^i, \ descending)$
10:     $\text{ALC}_i \leftarrow W^i[0 : m - 1]$
11:     Find n_c' in ALC_i by
      $\arg\max_{s \in \text{ALC}_i} \cos(C_{n_t}, C_{n_s})$
12:     $n_c \leftarrow n_c'$
13:     $LC.\text{enqueue}(n_c)$
14:    **end while**
15: **end while**

Alternative long chain Pruning. Then we designed alternative long chain pruning to select the

most important news among the i-th hop neighbors of n_t from ALC_i to add to the long chain, as the i-th news article for n_t's long chain. We calculate cosine similarity(C_{n_t}, C_{n_s}) for each s in ALC_i and select the most similar article s^* to add to the tail of long chain for n_t. Thus the long chain is refined based on semantics, category, and popularity to enhance the computation of similarities.

$$s^* = \arg \max_{s \in ALC_i} \cos(C_{n_t}, C_{n_s}) \qquad (7)$$

We iterate alternative long chain selection and pruning in different hop neighbor sets of n_t. Thus in each hop of n_t's neighbor, we compress the exponentially large neighbor set $N_{n_t}^{i+1}$ to a constant size $O(1)$.

In summary, we randomly initialize parameters of local news encoder for an initial long chain selection, and then periodically reselect after a certain number of training steps. During the training news representation process, we gradually extract semantic, category, and popularity information to enhance selection ability. This dynamic process ensures valuable long chains are identified during each reselection as shown in algorithm 1.

Loop chain. Long chain selection inherently includes recursive loops, such as $n_A \rightleftharpoons n_B$, resulting in sequences like $n_A \rightarrow n_B \rightarrow n_A \rightarrow n_B$. While avoiding loops could reduce redundancy, it may also limit the distillation of long chain interest. Therefore, we intentionally allow such cycles.

3.2.2 Long chain Interest Encoder

To effectively capture far-reaching linkage, we propose a long chain attention network to distill long chain interest.

At the news level, we now have the most valuable long chain for every clicked news, the correlation between different nodes in one long chain is typically similar news with a chronological progression. For instance, interest in football can be identified from several non-adjacent football news clicks in this long chain. Thus, we propose to apply news-level multi-head self-attention to enhance the the long chain interest representations l_n of news by capturing these non-adjacent interest. The corresponding formula is as follows:

$$\mathbf{l}_n = \sum_{i=1}^{len} \beta_i C_i \qquad (8)$$

$$\beta_i = \frac{exp(\mathbf{q}^\top \tanh(\mathbf{W}_1 \cdot \mathbf{C}_n^i))}{\sum_{j=1}^{len} exp(\mathbf{q}^\top \tanh(\mathbf{W}_1 \cdot \mathbf{C}_n^j))} \qquad (9)$$

where $\mathbf{q}$ is a query vector, $\mathbf{W}_1$ is a trainable matrix, C_n^i and C_n^j represent the category enhanced news representation by local news encoder, β_i represents the attention weight of a specific hop C_n^i in the long chain of the current news learned through multi-head self-attention, l_n represents the long chain representation of the current clicked news.

Different long chain may have different informativeness in representing users. Therefore, we personalize user interest on these long chains. At the user level, we propose to apply an additive attention network to select important long chain interest to learn more informative user representations. This allows each user to form its own specific long chain interest representation $\mathbf{L}_u$ based on $\mathbf{H_u}$.

$$\mathbf{L}_u = \sum_{i=1}^{L_{his}} \gamma_i \mathbf{l}_n^i \qquad (10)$$

$$\gamma_i = \frac{exp(\mathbf{q}^\top \tanh(\mathbf{W}_2 \cdot \mathbf{l}_n^i))}{\sum_{j=1}^{L_{his}} exp(\mathbf{q}^\top \tanh(\mathbf{W}_2 \cdot \mathbf{l}_n^j))} \qquad (11)$$

where $\mathbf{q}$ is a query vector, $\mathbf{W}_2$ is a trainable matrix. γ_i represents the attention weight of the long chain representation $\mathbf{l}_n$ of a clicked news article for the current user that we have learned.

3.3 Neighbor Interest Encoder

In the third phase of neighbor interest distillation, for each user click history, we construct a subgraph from the global click graph to analyze specific user interest. For each news in $\mathbf{H_u}$ of user u, we select the top M_n neighbors from multiple hops in the global click graph based on popularity in global click graph. We employ the gated GNN, which integrates a gated recurrent unit (Cho et al., 2014), to encode the neighbor news embedding:

$$\mathbf{N}_n = \text{GGNN}(\mathbf{h}_n^{ln}) \qquad (12)$$

$$\mathbf{h}_i^{(l+1)} = \text{GRU} \left(\sum_{j \in \mathcal{N}(i)} W_g \cdot \mathbf{h}_j^{(l)}, \mathbf{h}_i^{(l)} \right) \qquad (13)$$

where $\mathbf{h}^{(0)} = \mathbf{h}^{ln}$, initializing with $\mathbf{h}^{ln}$. $\mathcal{N}(i)$ represents the neighbors of node i, and W_g is a trainable matrix.

The neighbor aggregator, following (Wu et al., 2019b), utilizes a multi-head attention layer and an attention pooling layer to aggregate $\mathbf{O_u}$ to derive the user final long chain to enhance neighbor representation $\mathbf{LN}_u$, akin to Eq. 3.

3.4 Collaborative Interest Fusion and user representation

Considering the complexity of long chain interest and neighbor interest, we initially exclude using direct operations like multiplication, concatenation or addition. This is because these two interest are not entirely independent and often occasionally overlap. Therefore, we design a gated fusion network based on LSTM to balance long chain interest and neighbor interest with a common G_u to separately affect long chain and neighbor, thus model multi-level interest by integrating long chain interest L_u with current neighbor interest N_u:

$$\boldsymbol{G}_u = \text{sigmoid}(\boldsymbol{W}^1 \boldsymbol{N}_u + \boldsymbol{W}^2 \boldsymbol{L}_u + b) \quad (14)$$

Thus we can fuse long chain and neighbor interest to model collaborative interest $\mathbf{LN}_u$:

$$\mathbf{LN}_u = (\boldsymbol{1} - \boldsymbol{G}_u) \odot \boldsymbol{N}_u + \boldsymbol{G}_u \odot \boldsymbol{L}_u \quad (15)$$

where $\odot$ is element-wise multiplication, $\mathbf{W}^1$ and $\mathbf{W}^2$ are trainable matrices, b is a bias of fusion.

As for global-local aggregator, for each user, we derive the local news representation $\mathbf{h}^{ln}$, the local entity representation $\mathbf{h}^{le}$, and the long chain enhanced neighbor representation $\mathbf{LN}_u$. We employ an attention pooling network to combine these three representations into a user representation $\mathbf{emb}_{user}$.

3.5 Candidate News Modeling and Prediction

This candidate encoder leverages a global entity graph (Yang et al., 2023) for candidate news representation. The graph is constructed from entity occurrences in news. Entity encoding begins with selecting the top M_e neighbors for each entity in candidate news d_c. The local-view entity encoder's embedding layer is then used to derive $\mathbf{x}_c$, which is refined by $MSA(\mathbf{x}_c)$ to yield the global entity representation $\mathbf{h}_c^{ge}$. The candidate news representation $\mathbf{emb}_{cand}$ is synthesized by pooling the local news $\mathbf{h}_c^{ln}$, local entity $\mathbf{h}_c^{le}$, and global entity $\mathbf{h}_c^{ge}$ representations.

We calculate the click score for each news article by taking the inner product of the user embedding $\mathbf{emb}_{user}$ and the candidate embedding $\mathbf{emb}_{cand}$. We use negative sampling (Wu et al., 2019a) in training. In each session of the training set, we use one positive and K negative samples, whose click probability score is $\hat{y}^+$ and $\left[\hat{y}_1^-, \hat{y}_2^-, \ldots, \hat{y}_K^-\right]$ respectively. We then optimize the log-likelihood loss $\mathcal{L}_{NCE}$ for positive samples during training:

$$\mathcal{L} = -\sum_{i=1}^{|\mathcal{D}|} \log \left(\frac{e^{\hat{y}_i^+}}{e^{\hat{y}_i^+} + \sum_{j=1}^{K} e^{\hat{y}_{i,j}^-}} \right) \quad (16)$$

4 Experiment Setup

4.1 Dataset and Experiment Settings

Mind (Wu et al., 2020) is a real-world news recommendation corpus including Mind-small and Mind-large datasets, as shown in Table 3. We conduct our experiments on them. Following the previous work (Wu et al., 2019a), each user's historical record is constrained to include no more than $L_{his} = 50$ of the most recent news articles, with the length of each article's title not exceeding $L_{title} = 30$ words. We utilize up to $L_{entity} = 5$ entities per news article, adjacency entities involve 10 neighbor entities. The news neighbor interest encoder considers 2 hops with 10 neighbor news each hop. We employ a 2-layer GGNN (Li et al., 2016) for neighbor interest encoder. We initialize with 300-dimensional GloVe and 100-dimensional TransE embeddings from MIND dataset, configuring news representations at 400 dimensions. m is set to 3 to select top 3 ALC neighbors at each hop, while len is set to 8 for maximum length of a long chain. Adam (Kingma and Ba, 2015) optimizes our model with a learning rate of $2e^{-4}$, incorporating a 10% warm-up and linear decay. The negative sampling rate is set at $K_{neg} = 4$, with all settings validated against a test set.

4.2 Comparison Methods

We select four groups of baselines to compare: **Sequence-based methods**: **(1) LSTUR** (An et al., 2019a) combines GRU-based neighbor user interest embedding with user ID embedding for dynamic user representation. **(2) DAN** (Zhu et al., 2019) integrates LSTM attention and candidate awareness to dynamically model user interest from interaction history. **(3) CUPMAR** (Tran et al., 2021) uses GRU to understand user preferences from reading history, employing a news encoder for article properties and a user-profile encoder for user context interest.

Attention-based methods: **(1) NRMS** (Wu et al., 2019c) employs multi-head self-attention networks to derive representations for news and users. **(2) HieRec** (Qi et al., 2021b) utilizes a hierarchical interest tree to represent and match user interest at multiple granularities.

	MIND-small				MIND-large			
Method	AUC	MRR	nDCG@5	nDCG@10	AUC	MRR	nDCG@5	nDCG@10
CUPMAR	64.15	29.61	32.89	39.02	65.82	31.23	34.35	40.84
LSTUR	65.87	30.78	35.15	40.15	67.08	32.36	35.15	40.93
DAN	65.14	30.04	32.98	39.52	66.88	32.81	35.79	41.91
NRMS	65.36	30.02	33.11	39.61	67.01	31.85	35.34	41.75
HieRec	**67.95**	**32.87**	**36.36**	**42.53**	69.03	33.89	37.08	43.01
User-as-Graph	–	–	–	–	**69.23**	**34.14**	37.21	43.04
KIM [†]	67.07	31.83	35.23	41.58	68.45	33.74	36.76	42.47
DIGAT w/o PLM	67.82	32.65	36.25	42.49	–	–	–	–
GNewsRec	65.54	30.27	33.29	39.80	68.15	33.45	36.43	42.10
GLORY [†]	67.68	32.45	35.78	42.10	69.04	33.83	**37.53**	**43.69**
GLoCIM [†]	<u>68.21</u>	<u>33.02</u>	<u>36.69</u>	<u>42.78</u>	<u>69.52</u>	<u>34.34</u>	<u>37.89</u>	<u>44.08</u>

Table 2: Main results on MIND datasets. **Bold** means the best performance of baseline methods, while <u>**underline**</u> means GLoCIM performance, which is also the best in all metrics. We repeated these experiments with [†] three times and reported the average result. We outperform the best method at level p < 0.05.

	MIND-small	MIND-large
# News	65,238	161,013
# Categories	18	20
# Sub-categories	270	294
# Impressions	230,117	15,777,377
# Clicks	347,727	24,155,470

Table 3: Dataset Statistics.

GNN-based methods: **(1) User-as-Graph** (Wu et al., 2021) employs heterogeneous graph and graph pooling to depict complex user-action relationships. **(2) KIM** (Qi et al., 2021a) models user-news interactions using a co-encoder and a knowledge graph for user interest representation. **(3) DIGAT** (Mao et al., 2022) utilizes a dual graph attention network to align news and user graph channels for effective recommendation.

Global-view click graph methods: **(1) GNews-Rec** (Hu et al., 2020a) merges user and news graph data with topic categorization into a unified GNN model for enriched news representation. **(2) GLORY** (Yang et al., 2023) combines a global click graph with a gated GNN to create a hybrid news encoder that enhance global representation.

5 Experiment Results and Analysis

5.1 Main Comparison Results

Table 2 presents the performance comparison results, with GLoCIM consistently achieving the highest across all metrics. We select three sets of experiments for comparison.

First, on the Mind-small dataset, HieRec (Qi

et al., 2021b) outperforms DIGAT (Mao et al., 2022) by 0.17 in AUC, highlighting local user click history attention benefits in a smaller dataset.

Next, on Mind-large dataset, User-as-Graph outperforms HieRec by 0.2 in AUC, showing that GNN-based methods with higher-dimensional information excel over those based on user local click history attention in larger dataset.

Finally, on the same dataset, GLORY (Yang et al., 2023) achieves substantial gains in nDCG@5 and nDCG@10, increasing by 0.32 and 0.65 respectively, emphasizing the superiority of global click graph methods in enhancing ranking effectiveness.

Compared to the three aforementioned suboptimal methods, our method shows a clear advantage. In the case of Mind-small dataset where HieRec holds an advantage, GLoCIM's integration of long chain interest, combined with the use of the global click graph under data scarcity, enables it to outperform HieRec by a margin of 0.26 in AUC. On Mind-large dataset, GLoCIM also excels, surpassing the suboptimal User-as-Graph by 0.29 in AUC and GLORY by 0.36 in nDCG@5 and 0.39 in nDCG@10. This performance improvement comes from the fact that GLoCIM not only utilizes the neighbor interest but also distills long chain interest in the global click graph to capture far-reaching linkage and effectively fuse them to enhance user representation.

5.2 Ablation Analysis

To verify the validity of our fundamental modules, we conducted four sets of experiments to dive into the effectiveness of the Long chain Interest Encoder (LE), Long chain Selection Algorithm
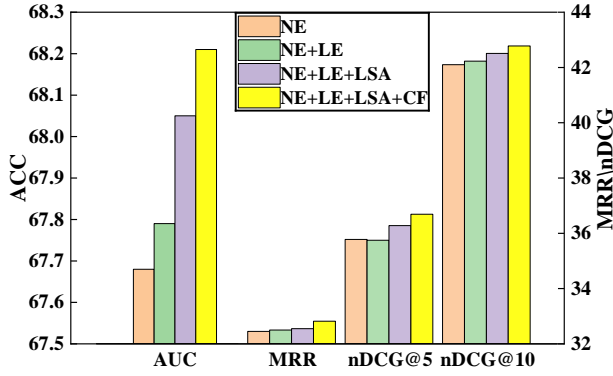
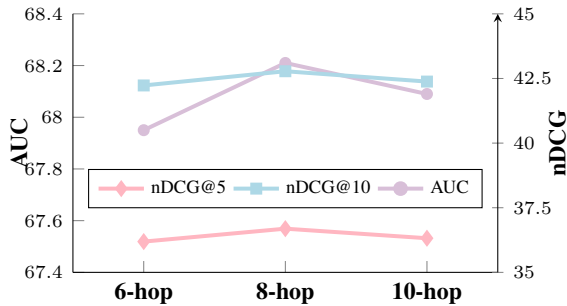Figure 3: Results of different module variants.

Figure 4: Impact of different hops

Model	AUC	Training Time (s)	Inferring Time (s)	GPU Mem (GB)
GGNN-1	67.35	10.62	11.80	6.84
GGNN-2	**67.68**	21.02	21.02	13.60
GGNN-3	67.49	121.91	55.84	20.43
GAT-1	67.40	10.85	12.62	7.65
GAT-2	67.60	21.78	23.55	14.37
GAT-3	67.58	113.91	59.05	18.82
GCN-1	67.22	9.94	16.70	12.39
GCN-2	67.45	21.01	15.09	17.32
GCN-3	67.31	110.09	52.49	20.43
GLoCM-8	**68.21**	57.85	24.89	14.97

Table 4: Model performance comparison. Average time cost of training/inferring 1000 samples.

(LSA), and Collaborative Interest Fusion Module (CF), with our base module being the Neighbor Interest Encoder (NE). We conducted four experiments: (1) **NE**: only use neighbor interest encoder (2) **NE+LE**: incorporate Long chain Interest Encoder but select long chain only by popularity (3) **NE+LE+LSA**: incorporate LSA but the fuse module is only element-wise multiplication (4) **NE+LE+LSA+CF(GLoCIM)**: full method.

All other settings remain the same as the full method, with only one component being replaced or removed. We evaluated these experiments on MIND-small dataset. The experimental results are shown in Figure 3. The full method performs the best, but the individual components also play a critical role in achieving optimal results. The role of LSA is the most apparent, proving that LSA selects the most valuable long chains, thereby effectively extracting and utilizing the long chain information in the global click graph.

5.3 Analysis on Hyperparameters

We explore the hyperparameter of hop number len in a long chain to assess its impact on performance. Experimental results from MIND-small dataset, as illustrated in Figure 4, indicate that GLoCIM at-

tains optimal recommendation performance when the number of hops is set to 8. This finding suggests that a lower number of hops may be inadequate for effectively distilling essential long chain interest, while a higher number may introduce excessive noise, thereby degrading the performance.

5.4 Impact of Graph Encoder

Far-reaching linkage modeling in global click graph faces massive computational overhead. So we evaluate impact of graph encoder on MIND-small dataset as shown in Table 4.

In order to extract far-reaching linkage, we adopt a common approach to compare: extracting local subgraphs from the global click graph to capture farther nodes, i.e., expanding the number of node hops. We use three popular GNNs (GAT, GCN, GGNN) to model the subgraphs. Observations show that as hops increase, training, inferring time, and overhead grow exponentially, while performance declines, indicating traditional methods' inefficiency in capturing far-reaching linkage.

Subsequently, our findings show that GLoCIM's AUC surpasses that of the best-performing GGNN-2 by over 0.53, indicating a more thorough exploitation of the global click graph and a better distillation of long chain interest that are often overlooked, while maintaining similar inference time and storage costs. It is important to note that our training time exceeds that of the optimal GGNN-2, which is reasonable given the time costs associated with the long chain selection algorithm. However, considering that the update period in the recommendation real-world scenario are not frequent, we believe that investing in longer training time to achieve higher accuracy is acceptable.
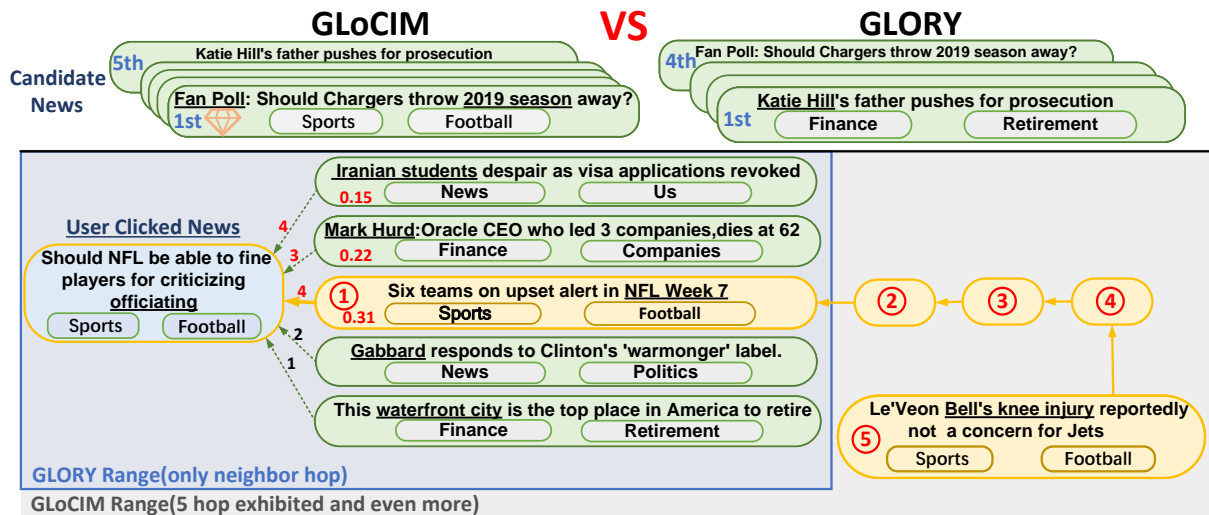
Figure 5: Comparison of the effect of GLoCIM and GLORY in utilizing global click graph for news recommendation. The circled red numbers represent different selected hop, and numbers on the arrows represent the accumulated frequency. The red numbers in the news represent the similarity to the user click news selected. The diamond icon represents the news that was actually clicked.

6 Case Study

We conduct a case study to evaluate GLoCIM and compare with GLORY, which is currently the best on the global click graph, as shown in Figure 5.

We display one clicked news article from a randomly chosen user in the same impression. We show the news suggested by GLoCIM and GLORY, where only one candidate news "Fan Poll" was clicked. We also present the selected representative long chain related to this clicked news. Several observations can be obtained from this.

Firstly, our long chain can select candidate news closely related to the news in the global click graph. our long chain selection includes "NFL Week 7" and "Le'Veon Bell's knee injury", rather than finance news such as "Mark Hurd" and "Waterfront City". We can clearly see our selection mechanism from the Figure 5. First, we calculate the corresponding weights based on accumulated frequency to obtain the alternative long chain: "Iranian students", "Mark Hurd" and "NFL Week 7". Then, by combining semantic and category information to obtain the category-enhanced representation, we calculate the cosine similarity and finally add the highest score "NFL Week 7" to our long chain. Repeating the above process, we can obtain the most valuable long chain. Therefore, although finance-related news appears twice in the first hop, it is still not considered as our candidate long chain.

Secondly, GLoCIM ranks candidate news clicked by the user higher than GLORY. We can observe that the current user actually clicked on "Fan Poll" in a real-world scenario, with GLoCIM recommending a rugby-related news article first, whereas GLORY first recommended "Katie Hill", a finance news article. This occurs as GLORY only uses GGNN to model user interest in neighbor hops. Since the neighbor news includes 'Mark Hurd' and 'waterfront city', GGNN over-learns news in the financial domain. Thus, GLORY ranks finance news higher. In contrast, we can not only capture neighbor news but also model relevant news representations in the most valuable long chain, such as the fifth hop news "Bell's knee injury", improving recommendation accuracy.

Different from GLORY, GLoCIM uses a long chain encoder and collaborative interest fusion, which can better distill long-chain interest in the global click graph than GLORY.

7 Conclusion

This paper presents GLoCIM, a novel news recommendation method that modeling a global-view long chain interest. We first design alternative long chain selection with news click popularity, then prune the alternative long chain based on enhanced news semantics and category representation. GLoCIM incorporates a neighbor and long chain interest encoder, distilling long chain and neighbor interest to enhance global click graph information extraction. Tests on two real-world datasets demonstrate that GLoCIM surpasses baseline methods.

8 Limitations

Training GLoCIM on static public news datasets with restricted time frames, which do not reflect the dynamic nature of real-world data. As a result, it may struggle to adapt to real-time changes in news recommendation, potentially impairing its generalization and robustness in evolving environments.

Acknowledgements

References

Mingxiao An, Fangzhao Wu, Chuhan Wu, Kun Zhang, Zheng Liu, and Xing Xie. 2019a. Neural news recommendation with long- and short-term user representations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 336–345, Florence, Italy. Association for Computational Linguistics.

Mingxiao An, Fangzhao Wu, Chuhan Wu, Kun Zhang, Zheng Liu, and Xing Xie. 2019b. Neural news recommendation with long-and short-term user representations. In *Proceedings of the 57th annual meeting of the association for computational linguistics*, pages 336–345.

Lei Chen, Le Wu, Richang Hong, Kun Zhang, and Meng Wang. 2020a. Revisiting graph based collaborative filtering: A linear residual graph convolutional network approach. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, New York, USA, February 7-12, 2020*, pages 27–34. AAAI Press.

Ming Chen, Zhewei Wei, Zengfeng Huang, Bolin Ding, and Yaliang Li. 2020b. Simple and deep graph convolutional networks. In *International conference on machine learning*, pages 1725–1735. PMLR.

Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, Doha, Qatar. Association for Computational Linguistics.

Abhinandan S Das, Mayur Datar, Ashutosh Garg, and Shyam Rajaram. 2007. Google news personalization: scalable online collaborative filtering. In *Proceedings of the 16th international conference on World Wide Web*, pages 271–280.

Shaohua Fan, Junxiong Zhu, Xiaotian Han, Chuan Shi, Linmei Hu, Biyu Ma, and Yongliang Li. 2019. Metapath-guided heterogeneous graph neural network for intent recommendation. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2478–2486.

Suyu Ge, Chuhan Wu, Fangzhao Wu, Tao Qi, and Yongfeng Huang. 2020. Graph enhanced representation learning for news recommendation. In *WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020*, pages 2863–2869. ACM / IW3C2.

Linmei Hu, Chen Li, Chuan Shi, Cheng Yang, and Chao Shao. 2020a. Graph neural news recommendation with long-term and short-term interest modeling. *Information Processing & Management*, 57(2):102142.

Linmei Hu, Siyong Xu, Chen Li, Cheng Yang, Chuan Shi, Nan Duan, Xing Xie, and Ming Zhou. 2020b. Graph neural news recommendation with unsupervised preference disentanglement. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4255–4264, Online. Association for Computational Linguistics.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Yujia Li, Richard Zemel, Marc Brockschmidt, and Daniel Tarlow. 2016. Gated graph sequence neural networks. In *Proceedings of ICLR'16*.

Zhiming Mao, Jian Li, Hongru Wang, Xingshan Zeng, and Kam-Fai Wong. 2022. DIGAT: Modeling news recommendation with dual-graph interaction. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 6595–6607, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Shumpei Okura, Yukihiro Tagami, Shingo Ono, and Akira Tajima. 2017. Embedding-based news recommendation for millions of users. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, page 1933–1942, New York, NY, USA. Association for Computing Machinery.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

Tao Qi, Fangzhao Wu, Chuhan Wu, and Yongfeng Huang. 2021a. Personalized news recommendation with knowledge-aware interactive matching. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 61–70.

Tao Qi, Fangzhao Wu, Chuhan Wu, Peiru Yang, Yang Yu, Xing Xie, and Yongfeng Huang. 2021b. HieRec: Hierarchical user interest modeling for personalized news recommendation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Online. Association for Computational Linguistics.

Dai Hoang Tran, Salma Hamad, Munazza Zaib, Abdulwahab Aljubairy, Quan Z Sheng, Wei Emma Zhang, Nguyen H Tran, and Nguyen Lu Dang Khoa. 2021. Deep news recommendation with contextual user profiling and multifaceted article representation. In *Web Information Systems Engineering–WISE 2021: 22nd International Conference on Web Information Systems Engineering, WISE 2021, Melbourne, VIC, Australia, October 26–29, 2021, Proceedings, Part II 22*, pages 237–251. Springer.

Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio, et al. 2017. Graph attention networks. *stat*, 1050(20):10–48550.

Hongwei Wang, Fuzheng Zhang, Xing Xie, and Minyi Guo. 2018. Dkn: Deep knowledge-aware network for news recommendation. In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, page 1835–1844, Republic and Canton of Geneva, CHE. International World Wide Web Conferences Steering Committee.

Chuhan Wu, Fangzhao Wu, Mingxiao An, Jianqiang Huang, Yongfeng Huang, and Xing Xie. 2019a. Npa: Neural news recommendation with personalized attention. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, page 2576–2584, New York, NY, USA. Association for Computing Machinery.

Chuhan Wu, Fangzhao Wu, Mingxiao An, Yongfeng Huang, and Xing Xie. 2019b. Neural news recommendation with topic-aware news representation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1154–1159, Florence, Italy. Association for Computational Linguistics.

Chuhan Wu, Fangzhao Wu, Suyu Ge, Tao Qi, Yongfeng Huang, and Xing Xie. 2019c. Neural news recommendation with multi-head self-attention. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6389–6394, Hong Kong, China. Association for Computational Linguistics.

Chuhan Wu, Fangzhao Wu, Yongfeng Huang, and Xing Xie. 2021. User-as-graph: User modeling with heterogeneous graph pooling for news recommendation. In *IJCAI*, pages 1624–1630.

Fangzhao Wu, Ying Qiao, Jiun-Hung Chen, Chuhan Wu, Tao Qi, Jianxun Lian, Danyang Liu, Xing Xie, Jianfeng Gao, Winnie Wu, and Ming Zhou. 2020. MIND: A large-scale dataset for news recommendation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3597–3606, Online. Association for Computational Linguistics.

Boming Yang, Dairui Liu, Toyotaro Suzumura, Ruihai Dong, and Irene Li. 2023. Going beyond local: Global graph-enhanced personalized news recommendations. In *Proceedings of the 17th ACM Conference on Recommender Systems*, pages 24–34.

Qiannan Zhu, Xiaofei Zhou, Zeliang Song, Jianlong Tan, and Li Guo. 2019. Dan: Deep attention neural network for news recommendation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 5973–5980.