

## Article

# FDDP-RN: Frequency-Domain Denoising and Popularity Bias Correction Recommendation Network

Xiaohui Du <sup>1,2,†</sup>, Yiwei Deng <sup>1,2,†</sup> , Xuelin Wang <sup>1,2</sup>, Biyang Ma <sup>1,2,\*</sup>  and Huifan Gao <sup>3,4,\*</sup>

<sup>1</sup> School of Computer Science, Minnan Normal University, Zhangzhou 363000, China

<sup>2</sup> Research Institute of Embodied Interaction Science and Technology, Minnan Normal University, Zhangzhou 363000, China

<sup>3</sup> Department of Automation, Xiamen University, Xiamen 361000, China

<sup>4</sup> Xiamen Airlines Co., Ltd., Xiamen 361006, China

\* Correspondence: mby@mnnu.edu.cn (B.M.); huifangao@stu.xmu.edu.cn (H.G.)

† These authors contributed equally to this work.

## Abstract

News recommendation is a critical technology that helps users efficiently find content of interest from large candidate pools. Its core objective is to accurately model user reading interests. However, current news recommendation systems typically suffer from two key limitations: (i) they fail to suppress noise from a frequency-domain perspective, and (ii) they lack effective calibration for popularity bias within the embedding space. In this work, we propose a novel frequency-domain denoising and popularity-bias correction recommendation network (FDDP-RN) to address both challenges simultaneously. Our approach introduces spectral analysis into the news encoder. Specifically, we design a filtering mechanism that combines truncation and scaling to enhance high-frequency semantic components, improve text feature representation accuracy, and suppress redundant low-frequency components. In addition, we introduce a norm-scaling factor that dynamically calibrates the embedding distribution of cold-start news items, placing them on an equal footing with popular news items. This effectively improves the exposure of long-tail content without requiring extra user interactions. We conduct extensive experiments on three public datasets, namely, MIND-small, MIND-large, and Adressa. The quantitative results demonstrate that FDDP-RN achieves state-of-the-art performance. Notably, on the Adressa dataset, our model achieves an AUC of 75.36% and an nDCG@10 of 50.11%, outperforming the strongest baseline. Furthermore, cold-start fairness diagnostics on the MIND-small dataset reveal that our mechanism increases the top-10 long-tail exposure rate from 15.3% to 18.1% and reduces the exposure Gini coefficient from 0.991 to 0.987, indicating a better balance among recommendation accuracy, diversity, and fairness.

**Keywords:** news recommendation; frequency-domain denoising; popularity; user interest modeling



Academic Editors: Sokratis Katsikas and Vincenzo Moscato

Received: 11 July 2026

Revised: 29 July 2026

Accepted: 30 July 2026

Published: 1 August 2026

**Copyright:** © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

## 1. Introduction

News recommendation serves as a core technology for modern content platforms. It directly affects user engagement and platform revenue. Poor recommendations cause users to lose interest, click less, and spend less time on the platform. Over time, this leads to user churn. Biased recommendations also reduce content diversity and fairness. Niche articles and socially important stories struggle to get attention. Therefore, improving both

accuracy and fairness in news recommendations is important for platform operators and content providers.

The main research problem is how to model user reading preferences from click history and news content. A good solution must handle two challenges. First, news text contains both stable information and sudden, time-sensitive signals. The stable part includes general topics and writing style. The sudden part includes breaking events, new entities, and changes in public attention. Second, popular news gets most of the clicks. This creates a feedback loop. Popular articles become more dominant in training data. Cold-start or long-tail articles get little exposure, even when they are relevant. A good recommender must be accurate, timely, and fair.

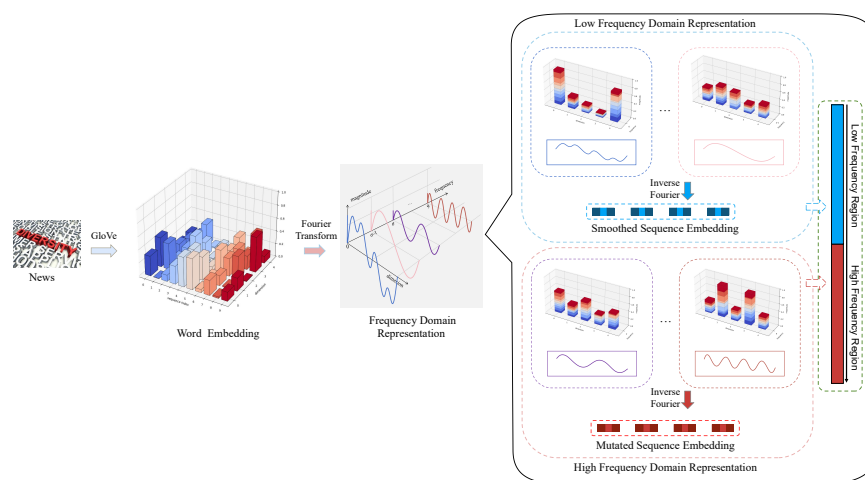
Many methods have been proposed for news recommendation. Early work used attention to extract features from user clicks. Zhu et al. [1] combined CNNs with attention and RNNs to model user interests. Qi et al. [2] designed a candidate-aware network that builds user representations for each candidate item. Kim et al. [3] used a context-aware network with a selection module to enrich headlines with body text. Recent work has explored graph-based models [4] and multi-interest learning [5]. These methods have pushed the field forward.

Despite this progress, existing methods have two key limitations. First, none of them leverage frequency-domain analysis for text encoding. As illustrated in Figure 1, news headlines inherently contain both low-frequency smooth semantics (e.g., general context) and high-frequency abrupt signals (e.g., named entities and event triggers) [6,7]. By stacking pretrained word vectors in token order and applying a fast Fourier transform (FFT) along the sequence dimension, our spectral analysis reveals how much each frequency contributes to the overall semantic meaning. Specifically, an empirical analysis of sampled news titles from the MIND-small dataset (Figure 1b) demonstrates that high-frequency energy accounts for between 35.51% and 48.27% of the total magnitude, with an average of 40.90%. This indicates that high-frequency peaks, driven by sudden keyword changes, are not negligible noise but a substantial part of the semantic content. Yet, existing attention models treat all words equally and inadvertently smooth out these high-frequency signals. This leads to recommendations that are safe but slow to reflect fresh content [8,9]. Second, existing methods lack effective calibration for popularity bias in the embedding space. Popular items dominate the training phase, pushing cold-start items to the tail of the distribution. This geometric bias persists in retrieval and ranking, making it difficult for fresh or niche content to surface even when it matches user interests.

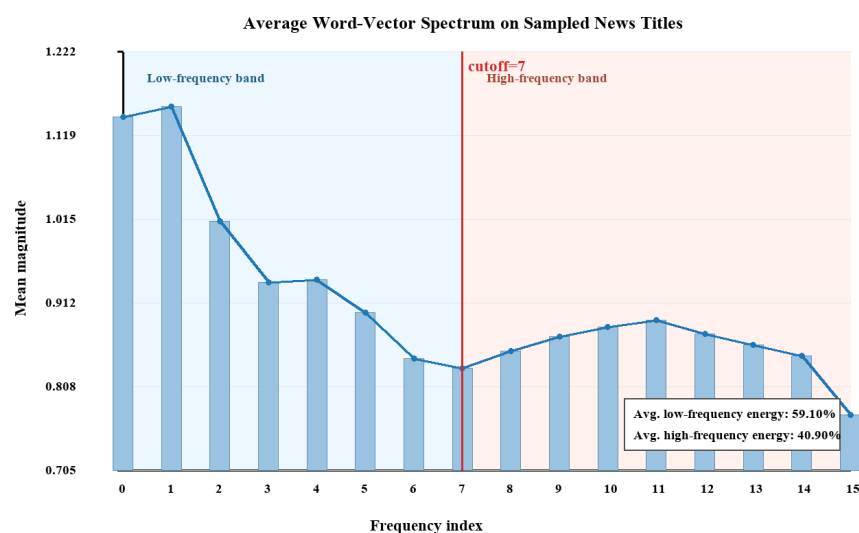
To determine whether the observed high-frequency components are discriminative rather than random noise, we further compute the per-frequency variance of magnitude spectra across news categories. Low-frequency components (index  $< 5$ ) have coefficients of variation between 0.12 and 0.18, whereas high-frequency components (index  $\geq 5$ ) exhibit  $1.8\times$  higher cross-category variance on average and up to  $2.3\times$  higher variance at peak frequencies. This contrast indicates that low-frequency components mainly preserve shared structural context, while high-frequency components contain stronger category-specific variation.

To address the above issues, we propose FDDP-RN, a network with frequency-domain denoising and popularity bias correction. The model has three main innovations. First, we apply FFT to word embeddings and design a truncation-and-scaling filter. This filter boosts high-frequency semantic components and reduces low-frequency noise. To the best of our knowledge, this is the first use of frequency-domain text processing in news recommendation. It addresses the problem that existing encoders smooth over sudden, meaningful signals. Second, we propose a click-response clustering method. It models popularity as a temporal pattern, not a fixed number. We extract features from each item's

click curve, such as growth rate, peak time, and decay slope. We then cluster items into popularity modes, such as cold-start or mainstream. Based on cluster membership, we apply a dynamic norm-scaling factor. This adjusts the embedding magnitude of cold-start items. It changes the geometry of the embedding space so that cold-start items are not at a disadvantage. To the best of our knowledge, this is among the first methods to perform popularity debiasing at the embedding level while supporting items with zero interactions. Third, we use an asymmetric routing design. Historically clicked news items go through the popularity calibration module. Candidate news items do not. This keeps candidate embeddings raw and rich. The asymmetry creates an implicit contrastive effect. The attention model learns to match on semantic direction, not embedding size. This separates relevance from popularity in the user model.



(a) Conceptual illustration of high-frequency and low-frequency signals.



(b) Empirical spectral analysis of word embeddings in sampled news titles.

**Figure 1.** Motivation for frequency-domain denoising. The combination of conceptual semantics (a) and empirical energy distribution (b) confirms that high-frequency abrupt signals constitute a substantial portion (average 40.90%) of the overall text features, which are often overlooked by traditional spatial-domain encoders. The colors in the schematic distinguish smooth low-frequency semantics from abrupt high-frequency signals; the plotted colors in the empirical panel distinguish the corresponding frequency components.

We test FDDP-RN on three public datasets: MIND-small, MIND-large, and Adressa. Results show that our model outperforms state-of-the-art baselines on all metrics and all datasets. We also conduct detailed ablations. For FDD, we compare against low-pass only, high-pass only, band-pass only, fixed scaling, learnable frequency-wise filtering, and no FDD. For PBC, we compare against popularity normalization, IPW, popularity regularization, and no PBC. These tests confirm that the gain comes from our specific designs, not from extra parameters.

Our main contributions are:

- We introduce a frequency-domain news encoder tailored to news recommendation. Unlike Fourier-based language models that mainly use spectral transforms for efficient token mixing, the proposed encoder performs explicit frequency partitioning on title word embeddings and reconstructs denoised news representations for recommendation.
- We design a truncation-and-scaling frequency denoising mechanism that selectively enhances high-frequency abrupt semantic signals while suppressing redundant low-frequency components, thereby improving the representation of timely and event-sensitive news.
- We propose an embedding-space popularity bias correction mechanism for cold-start news. Unlike click-count reweighting or popularity prediction methods, it calibrates the magnitude of cold-start news vectors toward a popularity-derived norm benchmark while preserving semantic direction.
- We conduct experiments on MIND and Adressa, together with mechanism-level spectral analysis, propagation-state clustering, ablation studies, diversity analysis, and cold-start fairness diagnostics, to verify both recommendation accuracy and the intended bias-correction behavior.

The rest of this paper is organized as follows. Section 2 reviews related work and gives background on FFT. Section 3 describes FDDP-RN in detail. Section 4 covers training and complexity. Section 5 presents the experimental setup. Section 6 shows results and analysis. Section 7 concludes and discusses limitations and future work.

## 2. Related Work

### 2.1. Personalized News Recommendation

Personalized news recommendation aims to help users find relevant articles from large-scale information streams. Early methods used collaborative filtering or content-based approaches. With the rise of deep learning, neural models became the dominant approach. Attention networks can effectively improve recommendation performance by extracting features from users' historical behavior. NRMS [10] used multi-head self-attention to learn news and user representations. NAML [11] extended this with multi-view attention on titles, bodies, and categories. CAUM [2] introduced candidate-aware user modeling that builds user representations conditioned on each candidate item. Wang et al. [12] detected the potential interest of each news article and assigned it to its corresponding interest-specific channel, after which a GRU-based network was employed to generate the personalized-interest representation for the session. Li et al. [5] further captured diverse user interests by learning multiple interest vectors through a poly attention scheme with disagreement regularization, and incorporated news category information via a category-aware attention weighting strategy to enhance personalized recommendation. Wang et al. [13] jointly modeled users' goal-oriented reading intentions and content-level reading preferences, aggregating both through a gate network for more accurate personalized news recommendations. More recent work has explored graph-based methods and multi-interest learning. Jiang et al. [4] proposed NRCLS, which enhances personalized recommendations by mod-

eling long-term preferences through a candidate-enhanced graph attention network and short-term preferences through a subtopic-aware candidate attention mechanism. These methods have improved accuracy but share a common limitation: they process text in the time or token domain and do not consider frequency-domain characteristics.

## 2.2. Frequency Domain and the Fourier Transform

The Discrete Fourier Transform (DFT) is a fundamental tool for analyzing the frequency-domain characteristics of discrete sequences. Its primary function is to decompose a discrete signal from the spatial domain into its spectral representation in the frequency domain, thereby enabling the analysis of the signal's spectral structure and patterns of variation. For a discrete input sequence  $f[l]$  with length  $L$  indexed by  $l = 0, 1, \dots, L - 1$ , its discrete Fourier transform  $F[k]$  is defined as:

$$F[k] = \sum_{l=0}^{L-1} f[l] e^{-j \frac{2\pi}{L} kl}, \quad k = 0, 1, \dots, L - 1 \quad (1)$$

Here,  $j$  denotes the imaginary unit, and  $F[k]$  represents the  $k$ -th spectral component in the frequency domain, comprising real and imaginary parts that determine the amplitude and phase information of the signal at that frequency, respectively.

Conversely, the Inverse Discrete Fourier Transform (IDFT) reconstructs the spatial domain sequence from its frequency-domain representation. Given a frequency-domain sequence  $F[k]$ , the corresponding spatial domain sequence  $f[l]$  is obtained as:

$$f[l] = \frac{1}{L} \sum_{k=0}^{L-1} F[k] e^{j \frac{2\pi}{L} kl}, \quad l = 0, 1, \dots, L - 1 \quad (2)$$

Compared to the DFT, the IDFT features a positive sign in the exponential term and incorporates a normalization factor of  $\frac{1}{L}$ . Through the IDFT, the model can perfectly reconstruct the modified spectral features back to the original data space, completing the signal reconstruction process.

Fourier transforms have been used in NLP for efficient sequence modeling. FNet [7] replaced self-attention with FFT for token mixing, reducing complexity. GFNet [14] used global filters in the frequency domain for image classification. In recommendation, SpectralCF [15] applied spectral analysis to user-item interaction matrices. More recently, FS-DRec [16] used frequency-sensitive diffusion for sequential recommendation. FITMM [17] applied frequency-aware learning to multimodal recommendation. These methods use frequency analysis on interactions or general features. None of them apply frequency-domain processing to news text content. To the best of our knowledge, our work is the first to do so.

## 2.3. Popularity Bias and Fairness

Unlike static property features typically exhibited after news release, news propagation follows a nonlinear saturation model [18]. As a measure of propagation, popularity reflects the attractiveness and freshness of news. For instance, Qi et al. [19] synthesized near-real-time CTR, recency, and title-based popularity prediction into a unified news popularity score. Meng et al. [20] used the time since news publishing and near-real-time CTR to model the current positions of news articles in their lifecycles. Ma et al. [21] estimated the ultimate popularity and propagation speed of news through a first-order saturation model, and employed a logarithmic time-decay mechanism to construct dynamic popularity scores so as to mitigate overexposure of trending content. The results indicate that these methods can capture the nature of news by incorporating additional popularity information. However, these methods tend to establish a relationship between click-through rates

and popularity, causing the models to overemphasize trending news and continuously promote highly visible content. Furthermore, Yang et al. [22] noted that popularity-based recommendations significantly influence users' news selection behavior; such choices often reflect the audience's preferences, leading to recommendations that favor popular content. While this improves the efficiency of information retrieval, it may also limit users' exposure to public affairs of social importance, thereby affecting the fairness of news dissemination.

Many methods have tried to fix popularity bias. PENR [23] added popularity prediction as an auxiliary task but kept item representations unchanged. Wang et al. [24] applied frequency-aware contrastive learning at the user modeling level. This did not calibrate cold-start item embeddings. PPSR [9] performed semantic retrieval for popularity prediction. It did not correct the geometric bias of cold-start items. Other approaches used inverse propensity weighting, causal inference, or regularization. These methods intervene at the loss, sampling, or ranking level. They do not change the embeddings themselves. Cold-start items remain geometrically disadvantaged in the representation space. Our method differs. We calibrate embedding magnitudes directly based on popularity patterns. This fixes the problem at its source.

#### 2.4. Summary

To summarize, existing work in news recommendation lacks frequency-domain text encoding. Frequency-domain methods in NLP and recommendation focus on efficiency or interaction data, not news text. Popularity debiasing methods work at the loss or ranking level, not at the embedding level. FDDP-RN fills all three gaps. It provides frequency-domain denoising for news text, embedding-level popularity calibration for cold-start items, and asymmetric routing that creates implicit contrastive dynamics.

### 3. Method

In this section, we present a frequency-domain denoising and popularity bias correction recommendation network (FDDP-RN) for news recommendation. As shown in Figure 2, FDDP-RN comprises two main components: frequency-domain denoising is introduced prior to the self-attention mechanism to amplify salient, high-frequency keywords in titles while attenuating smooth, redundant segments. This prevents semantic noise from diluting core signals, thereby enhancing the discriminability of news representations. Popularity bias correction is introduced to mitigate the unfairness faced by cold-start news due to a lack of user interaction signals. By dynamically scaling the embedding vectors of cold-start news, this mechanism aligns their distribution with the mean of popular samples, effectively suppressing the inherent marginalization of niche content.

#### 3.1. Problem Formulation and Notations

This section sets the formal framework for our news recommendation system, introducing key mathematical notations and conceptual foundations. To maintain notational consistency throughout our methodology, the primary symbols and their corresponding descriptions used in FDDP-RN are systematically summarized in Table 1.

We define and formalize the news recommendation task as follows: for each individual user  $u \in \mathcal{U}$ , their historical click sequence is denoted as  $\mathcal{N}^u = \{n_i^u\}_{i=1}^{|\mathcal{N}^u|}$ . Each historically clicked news article  $n_i^u$  is represented by a word sequence  $\mathcal{W}_{n_i^u} = \{w_j\}_{j=1}^L$ , where  $L$  is the length of the sequence. Given the user's historical interaction sequence  $\mathcal{N}^u$  and a candidate news item  $n_c \in \mathcal{N}_c$ , our ultimate objective is to predict the user's click probability  $\hat{y}_u$  for the candidate news item  $n_c$ .

**Table 1.** Key Notations and Descriptions for FDDP-RN.

| Symbol                                | Description                                          | Symbol                                                | Description                                           |
|---------------------------------------|------------------------------------------------------|-------------------------------------------------------|-------------------------------------------------------|
| $\mathcal{N}^u$                       | Historically clicked news sequence of user $u$       | $\mathcal{N}_c, n_c$                                  | The set of candidate news and a candidate item        |
| $\mathcal{W}_{n_i}$                   | The word sequence of news article $n_i$              | $\mathbf{X}$                                          | Spatial-domain word embedding matrix                  |
| $\hat{\mathbf{X}}, \hat{\mathbf{X}}'$ | Original and gated frequency representations         | $\sigma$                                              | Truncation index for low-frequency components         |
| $\zeta$                               | Scaling factor for high-frequency components         | $\tilde{\mathbf{X}}$                                  | Denoised spatial-domain word matrix                   |
| $\mathbf{r}_{n_i}$                    | Initial news representation generated by the encoder | $\eta_{n_i}^*, T_{n_i}^*$                             | Popularity ceiling and propagation saturation time    |
| $\mathbf{v}_{n_i}$                    | 2D state vector characterizing news popularity       | $\mathcal{N}^{\text{cold}}, \mathcal{N}^{\text{pop}}$ | Partitioned cold-start and popular news collections   |
| $\mu^{\text{pop}}$                    | Mean embedding magnitude of popular news collection  | $\zeta^{\text{cold}}$                                 | Popularity bias correction factor for cold-start news |
| $\varsigma$                           | Hyperparameter controlling the adjustment magnitude  | $\hat{\mathbf{x}}_{n_i}^u$                            | Calibrated news representation via PBC module         |
| $\kappa_j, \alpha_{n_i}^u$            | Word-level and news-level attention weights          | $\mathbf{W}_w, \mathbf{W}_n$                          | Projection matrices for dual attention                |
| $\mathbf{q}_w, \mathbf{q}_n$          | Query vectors for dual attention                     | $\mathbf{u}$                                          | Unified user interest representation                  |
| $\hat{y}_u$                           | Predicted click probability for candidate news       | $\Theta$                                              | The set of all learnable model parameters             |

### 3.2. Frequency-Domain Denoising

The news encoder aims to generate news representations with high distinctiveness. However, news titles often contain segments with gradual semantic changes and sparse information density, which appear as low-frequency components in the frequency domain. In traditional self-attention mechanisms, these segments tend to accumulate high weights due to their dominant sequence length, leading to the dilution of the core semantic signal. To address this, we propose a frequency-domain denoising strategy (FDD), which enhances the news encoder's ability to capture abrupt semantic changes by introducing a frequency-domain inductive bias.

For a generic news article  $n_i \in \mathcal{N}$  (representing either a historically clicked news item or a candidate news item), let  $\mathcal{W}_{n_i} = [w_1, \dots, w_L]$  denote its word sequence. The initial word embeddings  $\mathbf{x}_{w_j} = \text{GloVe}(w_j), \forall w_j \in \mathcal{W}_{n_i}$  [25] are stacked to form the embedding matrix  $\mathbf{X} = [\mathbf{x}_{w_1}, \dots, \mathbf{x}_{w_L}] \in \mathbb{R}^{L \times D}$ . Let  $K = \lfloor L/2 \rfloor + 1$  denote the number of non-redundant coefficients in the one-sided spectrum. Applying a real-valued fast Fourier transform along the sequence dimension gives:

$$\hat{\mathbf{X}} = \text{rFFT}(\mathbf{X}) \in \mathbb{C}^{K \times D}, \quad (3)$$

where rFFT stores only the non-redundant one-sided spectrum of the real-valued input. Because word embeddings are real-valued, the implementation uses a one-sided real FFT along the sequence dimension. Only the non-redundant coefficients are explicitly stored in the compact spectrum, while the negative-frequency coefficients are implicitly determined by conjugate symmetry. Filtering is applied in this compact domain, and the inverse real FFT is computed with the original sequence length specified so that the reconstructed word representation remains in  $\mathbb{R}^{L \times D}$ .

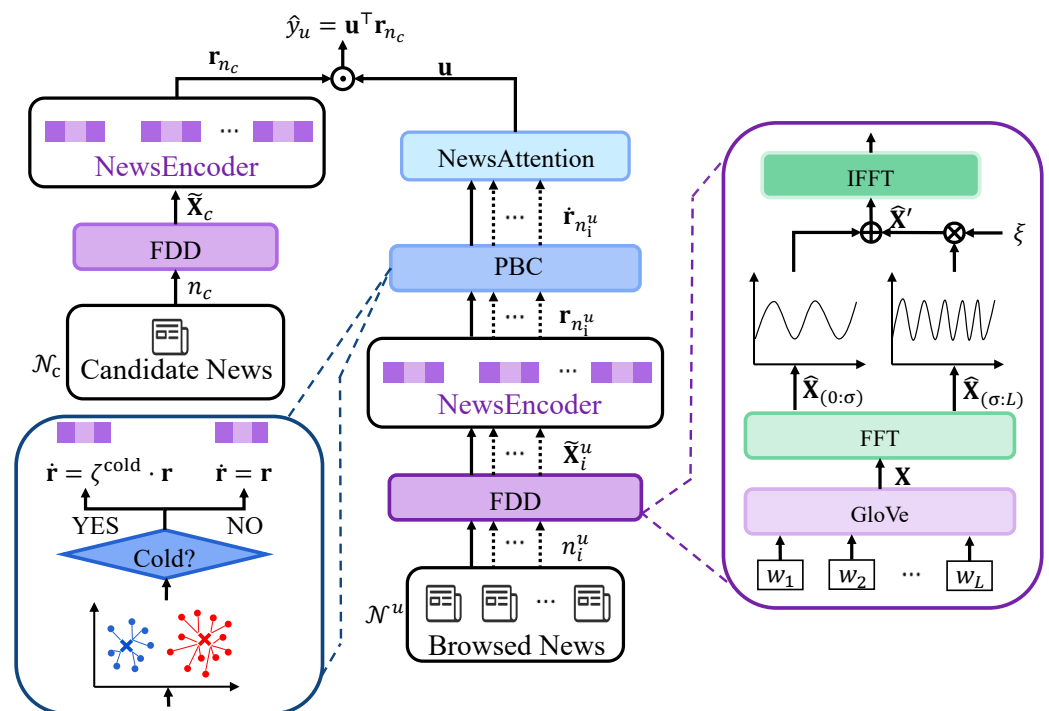


**Figure 2.** Overall architecture of FDDP-RN. The model follows the classic news-user encoding paradigm. News titles are first converted into word embedding sequences and transformed to the frequency domain via a fast Fourier transform (FFT). A learnable frequency-domain denoising mechanism suppresses low-frequency components and amplifies high-frequency components. The inverse FFT (IFFT) reconstructs the denoised embeddings, which are subsequently fed into a shared News Encoder to generate initial news representations. For candidate news, this representation  $r_{n_c}$  is directly utilized for click prediction. For historical interactions, the initial vectors are exclusively routed into a user interest modeling module, where a popularity bias correction (PBC) mechanism calibrates the magnitude of cold-start news. Finally, a news-level attention mechanism aggregates these calibrated vectors into a unified user interest representation  $u$ . The ultimate click probability is computed via the inner product. Colors distinguish the input, frequency-domain denoising, news encoding, popularity-bias correction, and user-modeling stages.

In the compact frequency representation  $\hat{X}$ , the first  $\sigma$  coefficients are treated as low-frequency components and the remaining  $K - \sigma$  coefficients as high-frequency components. We keep the low-frequency coefficients unchanged and introduce a learnable scaling factor  $\zeta \in \mathbb{R}$  for the high-frequency coefficients:

$$\hat{X}' = [\hat{X}'_k]_{k=0}^{K-1} \in \mathbb{C}^{K \times D}, \quad \text{where } \hat{X}'_k = \begin{cases} \hat{X}_k, & 0 \leq k < \sigma \\ \zeta \cdot \hat{X}_k, & \sigma \leq k < K \end{cases} \quad (4)$$

where  $\sigma$  is the truncation index determining the bandwidth retained for the low-frequency components. By adaptively strengthening key abrupt change signals,  $\zeta$  effectively enhances the sensitivity and discriminative power of news embeddings to important semantic variations. Accordingly,  $\sigma$  is the cutoff index in the compact one-sided spectrum. Coefficients below the cutoff are retained unchanged, while coefficients at and above the cutoff are multiplied directly by  $\zeta$ . The scaling coefficient  $\zeta$  is a learnable scalar optimized end-to-end via standard backpropagation together with the rest of the model. To keep the effective high-frequency amplification non-negative, we enforce a non-negativity constraint through a square or softplus reparameterization. The same scalar is shared across all samples, token positions, frequency bins in the high-frequency band, and embedding dimensions, so the

module introduces only a lightweight frequency-domain inductive bias rather than a large parameter matrix.

The modified compact spectrum is reconstructed with an inverse real FFT using the original sequence length:

$$\tilde{\mathbf{X}} = \text{irFFT}(\hat{\mathbf{X}}', n = L) \in \mathbb{R}^{L \times D}, \quad (5)$$

where  $\text{irFFT}(\cdot, n = L)$  restores the omitted negative-frequency coefficients through conjugate symmetry and reconstructs a real-valued sequence of length  $L$ .

### 3.3. News Encoder

To capture deep contextual dependencies, the reconstructed spatial-domain representation  $\tilde{\mathbf{X}} \in \mathbb{R}^{L \times D}$  from the FDD module is passed into the News Encoder. Built upon the architectural foundation of our prior work, PAD-MPFN [21], this module processes  $\tilde{\mathbf{X}}$  via a multi-head self-attention (MSA) [26] layer to capture global word interactions. Standard operations, including layer normalization, residual connections, and feed-forward networks (FFNs), are used in this module to enhance nonlinear representational capacity, yielding contextualized embeddings  $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_L]^\top$ :

$$\mathbf{H} = \text{MSA}(\tilde{\mathbf{X}}\mathbf{W}^Q, \tilde{\mathbf{X}}\mathbf{W}^K, \tilde{\mathbf{X}}\mathbf{W}^V) \in \mathbb{R}^{L \times D}, \quad (6)$$

where  $\{\mathbf{W}^Q, \mathbf{W}^K, \mathbf{W}^V\}$  are learnable projection matrices. Subsequently, a word-level additive attention network is applied to select informative words and aggregate them into the final news representation  $\mathbf{r}_{n_i} \in \mathbb{R}^D$ :

$$\mathbf{r}_{n_i} = \sum_{j=1}^L \kappa_j \mathbf{h}_j \in \mathbb{R}^D, \quad \kappa_j = \frac{\exp(\mathbf{q}_w^\top \tanh(\mathbf{W}_w \mathbf{h}_j))}{\sum_{k=1}^L \exp(\mathbf{q}_w^\top \tanh(\mathbf{W}_w \mathbf{h}_k))}, \quad (7)$$

where  $\mathbf{W}_w$  and the query vector  $\mathbf{q}_w$  are trainable parameters for word-level attention. The initial representation  $\mathbf{r}_{n_c}$  for the candidate news item  $n_c$  is generated via an identical encoding workflow, which explicitly bypasses the PBC module to maintain its objective semantic magnitude. For clarity in the algorithmic implementation (Algorithm 1), this entire word-level encoding process is abstracted as the function  $\text{NewsEncoder}(\cdot)$ .

### 3.4. Popularity Bias Correction

In news recommendation scenarios, traditional methods typically prioritize prediction accuracy as their core optimization objective. Because models based on inner products tend to assign larger feature magnitudes to popular items, they are prone to overfitting to trending news, resulting in an inherent popularity bias. This problem is particularly severe for newly released cold-start news: these news items not only lack real-world interaction data but also have initial embedding magnitudes that are often unevenly distributed. Specifically, some cold-start news items that accidentally acquire large initial magnitudes are incorrectly predicted as “high-popularity” and over-recommended, while most potentially relevant news items with smaller magnitudes are marginalized, severely violating the principle of fairness in recommendation.

To alleviate the above problems, we propose a popularity bias correction (PBC) mechanism. Following our prior work [21], we characterize the propagation dynamics of news  $n_i$  by its popularity upper bound  $\eta_{n_i}^*$  and propagation saturation time  $T_{n_i}^*$ . Specifically, given the cumulative click records  $\mathcal{D}_{n_i}^c(t) = \{(j, \tilde{C}_{n_i}(j))\}_{j=1}^t$ , we fit a first-order system response model:

$$C_{n_i}(t) = \eta_{n_i}(1 - e^{-t/T_{n_i}}), \quad (8)$$

by minimizing the following error:

$$\mathcal{L}_{\eta_{n_i}, T_{n_i}}(\mathcal{D}_{n_i}^c(t)) = \frac{1}{t} \sum_{j=1}^t (C_{n_i}(j) - \tilde{C}_{n_i}(j))^2. \quad (9)$$

---

**Algorithm 1** FDDP-RN training procedure
 

---

**Input:** Training set  $\mathcal{D}$ , truncation index  $\sigma$ , popularity upper bound  $\eta_{n_i}^*$ , propagation saturation time  $T_{n_i}^*$ , scaling factor  $\xi$ , negative sampling ratio  $K$ , batch size  $m$

**Output:** Learned model parameters  $\Theta$

```

1: Initialize word embedding  $\mathbf{X}$ ,  $\eta_{n_i}^*$ ,  $T_{n_i}^*$ , partitioning cold-start and popular news collec-
   tions  $\mathcal{N}^{\text{cold}}$  and  $\mathcal{N}^{\text{pop}}$  via Equation (12)
2: while epoch  $e = 1$  to  $E$  do
3:   Shuffle  $\mathcal{D}$  and partition into batches  $\{\mathcal{B}_1, \dots, \mathcal{B}_{|\mathcal{D}|/m}\}$ 
4:   for each batch  $\mathcal{B}$  do
5:     for each session  $\mathcal{S}^u \in \mathcal{B}$  do
6:        $n_i^+, \tilde{\mathcal{S}}_K^u \leftarrow \text{NegativeSample}(\mathcal{S}^u, K)$  ▷ Negative Sampling
7:       // Phase 1: Shared News Encoding (FDD & Context)
8:       for each news item  $n \in \{\mathcal{N}^u \cup \{n_i^+\} \cup \tilde{\mathcal{S}}_K^u\}$  do
9:          $\hat{\mathbf{X}} \leftarrow \text{rFFT}(\mathbf{X}_n)$  via Equation (3) ▷ One-sided real FFT
10:         $\hat{\mathbf{X}}' \leftarrow [\hat{\mathbf{X}}_k]_{k=0}^{L-1}$  via Equation (4) ▷ Frequency-domain Gating
11:         $\tilde{\mathbf{X}} \leftarrow \text{irFFT}(\hat{\mathbf{X}}', n = L)$  via Equation (5) ▷ Inverse real FFT
12:         $\mathbf{r}_n \leftarrow \text{NewsEncoder}(\tilde{\mathbf{X}})$  via Equation (7) ▷ Shared News Encoder
13:      end for
14:      // Phase 2: User Interest Modeling
15:      for each historical news item  $n_i^u \in \mathcal{N}^u$  do
16:         $\hat{\mathbf{r}}_{n_i^u} \leftarrow \text{PBC}(\mathbf{r}_{n_i^u})$  via Equation (15) ▷ Popularity Bias Correction
17:      end for
18:       $\mathbf{u} \leftarrow \text{NewsAttention}(\{\hat{\mathbf{r}}_{n_i^u} \mid n_i^u \in \mathcal{N}^u\})$  via Equation (16) ▷ User Interest Aggregation
19:      // Phase 3: Click Prediction
20:       $\hat{y}_i^+ \leftarrow \mathbf{u}^\top \mathbf{r}_{n_i^+}$  via Equation (17) ▷ Positive Sample Prediction
21:      for each  $n_{i,j}^- \in \tilde{\mathcal{S}}_K^u$  do
22:         $\hat{y}_{i,j}^- \leftarrow \mathbf{u}^\top \mathbf{r}_{n_{i,j}^-}$  via Equation (17) ▷ Negative Sample Prediction
23:      end for
24:    end for
25:    Compute batch loss  $\mathcal{L}$  via Equation (19) ▷ Loss Calculation
26:    Update  $\Theta$  via backpropagation
27:  end for
28: end while
29: return  $\Theta$ 

```

---

In implementation, the propagation curve is estimated only from the training split. For timestamped behavior logs, we first sort the training impressions by time and build an item-level cumulative positive-click curve. The popularity ceiling  $\eta_{n_i}^*$  is taken from the fitted asymptotic response value, while  $T_{n_i}^*$  denotes the first time point at which the cumulative curve reaches 63.2% of  $\eta_{n_i}^*$ . For the JSONL-format MIND-small analysis used in the reviewer-response scripts, where absolute timestamps are not retained, the ordered training records are divided into 48 chronological bins and the same 63.2% rule is applied to the binned curve. Thus, the PBC partition uses only pre-evaluation training evidence and never touches validation or test interactions. The first-order response parameters are fitted by nonlinear least squares using training data only. Direct item-level fitting requires at least three non-zero time points and five total positive clicks. Items below either threshold, including zero-interaction items, receive category-level mean estimates

computed exclusively from the training split. This fallback ensures that every item can enter the same propagation-state clustering pipeline without using validation or test evidence.

The optimized parameters  $\eta_{n_i}^*$  and  $T_{n_i}^*$  represent the popularity upper bound and propagation speed (time to reach 63.2% of saturation), respectively. These are projected into a two-dimensional feature space to construct the state vector:

$$\mathbf{v}_{n_i} = \left( \ln(1 + \eta_{n_i}^*), \ln(1 + 1/T_{n_i}^*) \right)^\top \in \mathbb{R}^2, \quad (10)$$

where the first dimension represents the popularity limit of a news item, and the second dimension indicates its spreading speed. Unlike traditional static click counts, this two-dimensional space can easily separate fast-trending news from long-term engaging content by capturing how they spread.

The clustering procedure is performed on news-level state vectors rather than on raw click counts or text tokens. Before clustering, the two coordinates in  $\mathbf{v}_{n_i}$  are log-transformed as shown above, which reduces the dominance of extremely popular articles and makes propagation ceiling and spreading speed comparable in scale. Each news article is therefore represented as one point in a two-dimensional propagation space, where the horizontal direction captures potential popularity capacity, and the vertical direction captures response speed. We then apply a bisecting K-means procedure with  $k = 2$  on the training-derived vectors. Compared with standard K-means, the bisecting variant is less sensitive to initialization and provides a clearer binary split between hot and cold content. The resulting partition is interpreted operationally: the cluster whose centroid lies closer to the origin is assigned to the cold-start collection  $\mathcal{N}^{\text{cold}}$ , while the other cluster is assigned to the popular collection  $\mathcal{N}^{\text{pop}}$ . This definition is more suitable for news recommendation than a rigid zero-click rule, because short-lived articles can still behave like cold-start items even after receiving sparse interactions. In the PBC stage, the cluster label is used only to decide whether magnitude calibration is applied; the semantic direction of the news embedding remains unchanged.

Starting with an initial cluster containing all state vectors, the algorithm iteratively performs binary splitting guided by the objective of minimizing the sum of squared errors (SSE) within clusters:

$$\text{SSE} = \sum_{\mathbf{v}_{n_i} \in \mathcal{V}} \|\mathbf{v}_{n_i} - \mathbf{c}_0\|^2, \quad \mathbf{c}_0 = \frac{1}{|\mathcal{V}|} \sum_{i=1}^{|\mathcal{V}|} \mathbf{v}_{n_i} \in \mathbb{R}^2, \quad (11)$$

where  $\mathbf{c}_0$  represents the centroid of the initial global cluster, and the resultant descendant cluster centroids are denoted as  $\mathbf{c}_k$  ( $k = 1, 2$ ) after the partition.

The cold-start and popular news sets are automatically identified based on the proximity of  $\mathbf{c}_k$  to the origin. Specifically, the cluster with a shorter distance to the origin is defined as the cold-start news collection  $\mathcal{N}^{\text{cold}}$ , whereas the one with a longer distance constitutes the popular news collection  $\mathcal{N}^{\text{pop}}$ :

$$\begin{cases} \mathcal{N}^{\text{cold}} = \{n_i \mid \mathbf{v}_{n_i} \in \mathcal{V}_k, \|\mathbf{c}_k\| = \min(\|\mathbf{c}_1\|, \|\mathbf{c}_2\|)\}, \\ \mathcal{N}^{\text{pop}} = \{n_i \mid \mathbf{v}_{n_i} \in \mathcal{V}_k, \|\mathbf{c}_k\| = \max(\|\mathbf{c}_1\|, \|\mathbf{c}_2\|)\}, \end{cases} \quad (12)$$

where the cold-start news set  $\mathcal{N}^{\text{cold}}$  corresponds to clusters whose centers are close to the origin, indicating that their features maintain low values in both the potential capacity and burst rate dimensions. The popular news set  $\mathcal{N}^{\text{pop}}$  corresponds to clusters whose centers are far from the origin, indicating that the news in this set has a higher propagation ceiling or a faster response speed. This partition should be understood as an operational cold-start

definition: a news item is treated as cold-start when its learned propagation state shows both limited expected reach and weak early response, even if it has already received a small number of clicks. This is more suitable for news recommendation than a rigid zero-click definition because news articles have short lifecycles and may remain under-exposed after receiving only sparse interactions. In the following PBC step, the cluster label is used only to decide whether magnitude calibration is applied; the semantic direction of the news embedding remains unchanged.

As shown in Figure 3 and Table 2, the fitted cold-start cluster contains 5495 news articles, accounting for 72.44% of the positive-click news pool, whereas the popular cluster contains 2091 articles, accounting for 27.56%. The median empirical popularity ceiling of the cold-start cluster is only 1.00, much smaller than that of the popular cluster (39.00), confirming that the proposed propagation-state partition identifies genuinely sparse-interaction news rather than arbitrarily splitting the item set. Using fewer than 10 training interactions as the conventional interaction-sparsity criterion, 85% of those low-interaction items are assigned to the propagation-based cold-start cluster. The proposed partition therefore largely overlaps with the conventional definition while additionally accounting for propagation ceiling and response speed.

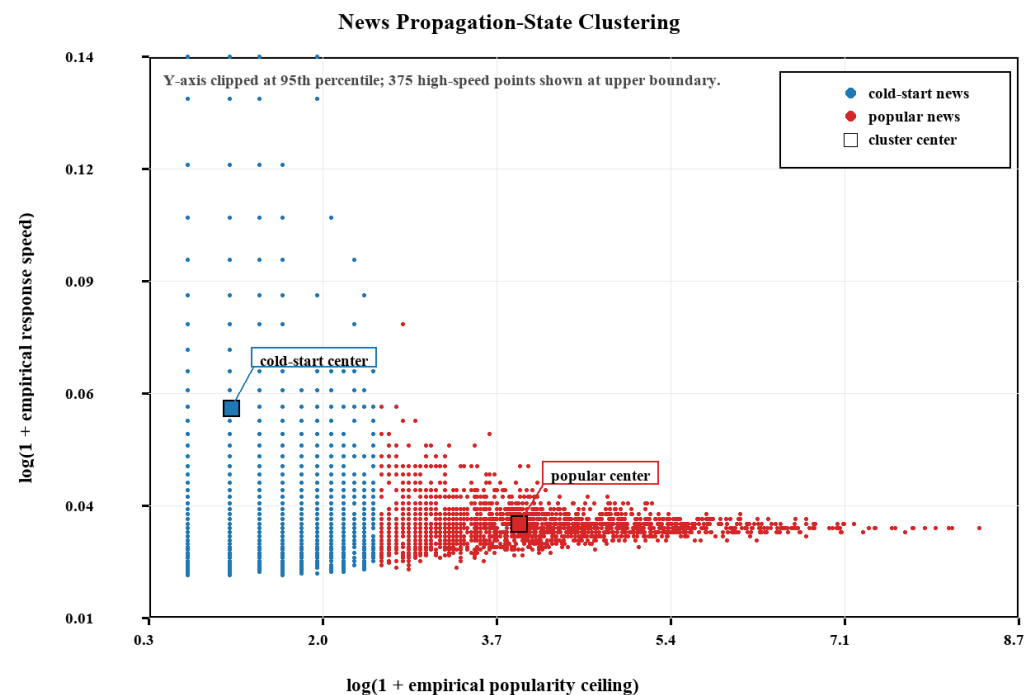


**Figure 3.** Two-dimensional propagation-state clustering of news articles on the MIND-small training split. Each point denotes one news article represented by  $(\ln(1 + \eta_{n_i}^*), \ln(1 + 1/T_{n_i}^*))$ . The cluster closer to the origin is identified as cold-start news, while the cluster farther from the origin is identified as popular news. For readability, the vertical axis is displayed up to the 95th percentile, and the few higher-speed points are shown at the upper boundary; the cluster statistics in Table 2 are computed on the full data.

**Table 2.** Statistics of propagation-state clusters on the MIND-small training split.

| Cluster    | News Count | Ratio  | Mean $\eta^*$ | Median $\eta^*$ | Center-x | Center-y |
|------------|------------|--------|---------------|-----------------|----------|----------|
| Cold-start | 5495       | 72.44% | 2.58          | 1.00            | 1.114    | 0.060    |
| Popular    | 2091       | 27.56% | 103.93        | 39.00           | 3.903    | 0.033    |

To balance the dominant influence of popularity on recommendation outcomes, we perform bias correction by calibrating the magnitude distribution of cold-start news embed-

ding vectors. Based on the initial news representation  $\mathbf{r}_{n_i}$  generated by the news encoder, the mean magnitude of the popular news collection  $\mathcal{N}^{\text{pop}}$  is calculated as a benchmark:

$$\mu^{\text{pop}} = \frac{1}{|\mathcal{N}^{\text{pop}}|} \sum_{n_j \in \mathcal{N}^{\text{pop}}} \|\mathbf{r}_{n_j}\| \in \mathbb{R}. \quad (13)$$

For a historically clicked cold-start news item  $n_i^u \in \mathcal{N}^{\text{cold}}$ , its correction factor  $\zeta^{\text{cold}}$  is defined to enable a linear contraction or expansion of the cold-start news magnitude distribution:

$$\zeta^{\text{cold}} = \frac{\|\mathbf{r}_{n_i^u}\| + \zeta \mu^{\text{pop}}}{\|\mathbf{r}_{n_i^u}\| (1 + \zeta)} \in \mathbb{R}, \quad (14)$$

where  $\zeta > 0$  is a tunable hyperparameter that determines the magnitude of the adjustment.

Ultimately, for any historically clicked news item  $n_i^u \in \mathcal{N}^u$ , we feed its initial vector  $\mathbf{r}_{n_i^u}$  into the PBC module to obtain the calibrated final representation  $\hat{\mathbf{r}}_{n_i^u}$  through a conditional operation:

$$\hat{\mathbf{r}}_{n_i^u} = \begin{cases} \zeta^{\text{cold}} \mathbf{r}_{n_i^u}, & \text{if } n_i^u \in \mathcal{N}^{\text{cold}} \\ \mathbf{r}_{n_i^u}, & \text{if } n_i^u \in \mathcal{N}^{\text{pop}} \end{cases} \in \mathbb{R}^D. \quad (15)$$

Through this calibration, cold-start news items acquire a calibrated magnitude distribution without altering their semantic direction.

The effect of Equation (15) can be inspected through the before/after norm distribution of cold-start and popular news vectors. As a mechanism-level diagnostic on the MIND-small training split, Figure 4 computes title-level vectors from averaged GloVe embeddings and applies the same PBC norm transformation with  $\zeta = 0.5$ . The cold-start mean norm moves from 3.277 before PBC to 3.282 after PBC, closer to the popular-news benchmark of 3.293, while the popular-news distribution is kept unchanged. This confirms that PBC performs magnitude calibration rather than semantic-direction rewriting.

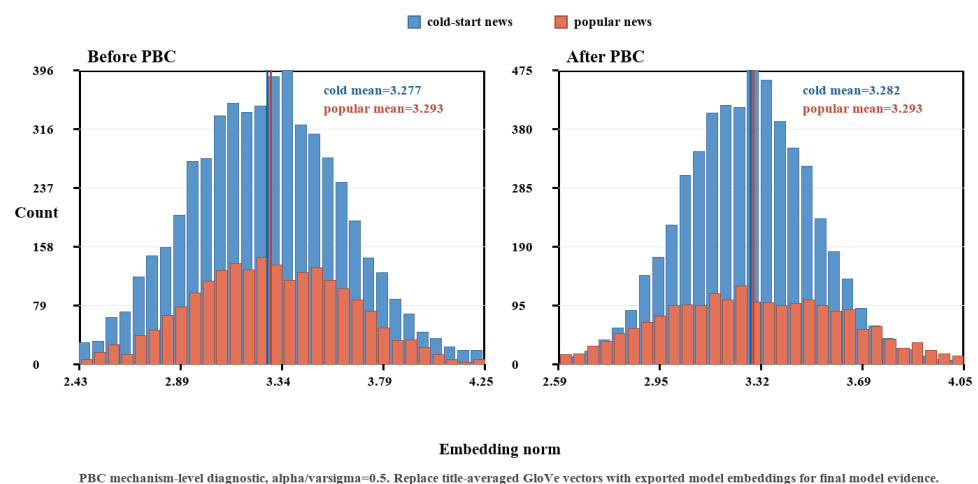


**Figure 4.** Before/after norm distribution under the PBC mechanism on the MIND-small training split. The default diagnostic uses title-averaged GloVe vectors and the same calibration formula as Equation (15); popular news vectors are unchanged, while cold-start norms are shifted toward the popular-news norm benchmark.

### 3.5. User Interest Aggregation

After being processed by the news encoder and the popularity bias correction, each news item  $n_i^u$  in the user's historical click sequence  $\mathcal{N}^u$  is mapped to its calibrated repre-

sensation  $\mathbf{r}_{n_i^u}$  via Equation (15). An attention mechanism is then employed to aggregate these individual interaction representations into a unified user interest vector  $\mathbf{u} \in \mathbb{R}^D$ :

$$\mathbf{u} = \sum_{n_i^u \in \mathcal{N}^u} \alpha_{n_i^u} \mathbf{r}_{n_i^u} \in \mathbb{R}^D, \quad \alpha_{n_i^u} = \frac{\exp(\mathbf{q}_n^\top \tanh(\mathbf{W}_n \mathbf{r}_{n_i^u}))}{\sum_{n_j^u \in \mathcal{N}^u} \exp(\mathbf{q}_n^\top \tanh(\mathbf{W}_n \mathbf{r}_{n_j^u}))}, \quad (16)$$

where  $\alpha_{n_i^u}$  is the sequential attention weight, and  $\mathbf{W}_n, \mathbf{q}_n$  are trainable parameters for news-level attention. The attention weight evaluates the informative value of each historically clicked news item in modeling the user's current preferences. For clarity in the algorithmic implementation (Algorithm 1), this news-level aggregation process is abstracted as  $\mathbf{u} = \text{NewsAttention}(\{\mathbf{r}_{n_i^u} \mid n_i^u \in \mathcal{N}^u\})$ .

### 3.6. Click Predictor

The click prediction module models the interaction between the user interest vector and the candidate news representation. For the candidate news item  $n_c$ , its representation  $\mathbf{r}_{n_c}$  is directly derived through the frequency-domain denoising (FDD) strategy and the identical NewsEncoder workflow from its word sequence  $\mathcal{W}_{n_c}$ .

Ultimately, the predicted click probability  $\hat{y}_u$  is computed as the inner product of the aggregated user interest embedding  $\mathbf{u}$  and the candidate news representation  $\mathbf{r}_{n_c}$ :

$$\hat{y}_u = \mathbf{u}^\top \mathbf{r}_{n_c} \in \mathbb{R}, \quad (17)$$

where  $\mathbf{u}$  incorporates the popularity bias correction derived from historical interactions, whereas  $\mathbf{r}_{n_c}$  explicitly bypasses the PBC module to preserve its objective semantic magnitude.

## 4. Model Training of FDDP-RN

Following standard training principles in news recommendation, we introduce a negative sampling mechanism during the model optimization stage to alleviate the imbalance between positive and negative sample distributions. It is worth noting that while our overall architecture is designed to score any candidate news  $n_c \in \mathcal{N}_c$ , during the training phase (Algorithm 1), we instantiate the candidate set through this negative sampling. Specifically, a batch  $\mathcal{B}$  containing  $|\mathcal{B}|$  independent user sessions is extracted from the training set  $\mathcal{D}$ . For each user session  $\mathcal{S}^u$ , we construct a localized candidate set consisting of one positive clicked item  $n_i^+$  and  $K$  randomly sampled negative items  $\mathcal{S}_K^-$  from the unclicked candidates. To accurately characterize user interests, we use the negative log-likelihood function as the optimization objective, which is defined as follows:

$$\hat{\mathbf{y}}_i = [\hat{y}_i^+, \hat{y}_{i,1}^-, \hat{y}_{i,2}^-, \dots, \hat{y}_{i,K}^-] \in \mathbb{R}^{K+1}, \quad (18)$$

where  $\hat{y}_i^+$  represents the predicted score of the positive sample, and  $\{\hat{y}_{i,1}^-, \dots, \hat{y}_{i,K}^-\}$  denote the predicted scores of the  $K$  corresponding negative samples in the  $i$ -th session. We optimize the model parameters by minimizing the batch-wise negative log-likelihood loss, which is formulated as:

$$\mathcal{L} = -\frac{1}{|\mathcal{B}|} \sum_{i=1}^{|\mathcal{B}|} \log \frac{\exp(\hat{y}_i^+)}{\exp(\hat{y}_i^+) + \sum_{j=1}^K \exp(\hat{y}_{i,j}^-)} \in \mathbb{R}, \quad (19)$$

where  $\mathcal{B}$  denotes the training mini-batch, and  $|\mathcal{B}|$  is the batch size.

This loss function aims to increase the scores assigned to positive samples and suppress the interference of negative samples by widening the prediction difference between positive and negative samples, thereby optimizing the model's performance in discriminating click

behavior. This allows both news representation and user representation to fully utilize frequency-domain information and popularity-correction information.

The FDDP-RN training procedure is detailed in Algorithm 1. Initially, word embeddings, popularity parameters, and state vectors are initialized to partition news into the cold-start news collection  $\mathcal{N}^{\text{cold}}$  and the popular news collection  $\mathcal{N}^{\text{pop}}$  (line 1). During each epoch (line 2), the shuffled data is optimized in batches. For each session  $\mathcal{S}^u$ , negative sampling constructs training instances (line 6). The model then computes representations following a structured three-phase process. In Phase 1: shared news encoding (lines 7–13), all involved news items (both historical and candidates) undergo frequency-domain denoising (FDD) and are encoded into initial representations  $\mathbf{r}_n$ . Subsequently, in Phase 2: user interest modeling (lines 14–18), the representations of historical clicked news items are exclusively routed to the popularity bias correction (PBC) module (line 16) to automatically generate fair representations  $\hat{\mathbf{r}}_n$  for both cold and popular news without altering semantic directions. A news-level attention mechanism then aggregates these calibrated vectors into a comprehensive user interest representation  $\mathbf{u}$  (line 18). Candidate news explicitly bypass this PBC module. Finally, in Phase 3: click prediction (lines 19–23), matching scores are calculated via inner products between  $\mathbf{u}$  and candidate representations (lines 20–22). The model is optimized by updating all parameters  $\Theta$  via backpropagation using the computed batch loss  $\mathcal{L}$  (lines 25–26).

The computational requirements of FDDP-RN can be analyzed through both theoretical complexity bounds and empirical measurements. (1) The encoding of a single news article undergoes a Fourier transform. An FFT is performed along the sequence dimension on the  $L \times D$  word vector matrix  $\mathbf{X}$  to transform the semantic signal from the spatial domain to the frequency domain. The computational cost is  $\mathcal{O}(L \cdot D \log L)$ . Scaling is performed through a frequency-domain denoising mechanism that splits high and low frequencies and scales high frequencies. This process involves only element-wise operations and no complex matrix operations. The computational complexity is  $\mathcal{O}(L \cdot D)$ . The computational cost of IFFT is the same as that of FFT, which is  $\mathcal{O}(L \cdot D \log L)$ . (2) Popularity bias correction only performs magnitude calibration and distribution normalization on the final embedding vector of historical cold-start news articles. No complex matrix operations or feature reconstruction are required. The calculation mainly includes estimating the average embedding magnitude of popular news, computing the embedding norm of cold-start news, and applying vector scaling based on the correction factor  $\zeta^{\text{cold}}$ . All operations are scalar operations and element-wise vector operations, with a computational complexity of  $\mathcal{O}(D)$ .

The computational feasibility of FDDP-RN is validated through empirical benchmarks on an NVIDIA GeForce RTX 4070 Ti SUPER GPU. During the training phase, the model achieves a processing throughput of 1.2 s per batch. This efficiency stems from the inherent  $\mathcal{O}(L \log L)$  complexity of the frequency-domain denoising module, which introduces minimal computational overhead compared to the standard spatial attention mechanisms. Additionally, the popularity bias correction operates as a parameter-free adjustment, ensuring that fairness calibration does not impede training scalability. Consequently, for a dataset with 250,000 samples, the model completes five full training epochs and reaches optimal convergence within approximately 10 h, demonstrating a superior trade-off between structural complexity and training speed. To quantify the additional cost introduced by the two proposed components, Table 3 reports the measured batch-level overhead under the same hardware and training configuration. Compared with the base news encoder without FDD or PBC, the full FDDP-RN model increases the average training time from 1.09 s to 1.20 s per batch, corresponding to a 10.1% overhead. The FDD module accounts for most of this increase because it performs one rFFT, one lightweight frequency scaling operation,

and one inverse real FFT along the title sequence dimension. The PBC module adds only 0.02 s per batch because it uses norm computation and scalar rescaling on historical news embeddings. It introduces no trainable parameters and requires only a one-time offline propagation-state clustering step.

**Table 3.** Computational overhead of the proposed FDD and PBC components. Runtime is measured on an NVIDIA GeForce RTX 4070 Ti SUPER GPU (NVIDIA Corporation, Santa Clara, CA, USA) with the same batch size and training protocol used in the main experiments.

| Setting      | Additional Operations                           | Time/Batch | Overhead | Extra Parameters |
|--------------|-------------------------------------------------|------------|----------|------------------|
| Base encoder | MSA-based news encoding only                    | 1.09 s     | –        | –                |
| +FDD         | rFFT, frequency scaling, and irFFT              | 1.18 s     | +8.3%    | 1 scalar         |
| +PBC         | Embedding-norm computation and scalar rescaling | 1.11 s     | +1.8%    | 0                |
| FDDP-RN      | FDD + PBC                                       | 1.20 s     | +10.1%   | 1 scalar         |

In terms of deployment and inference, if news embeddings are pre-generated and stored in a cache, the computational overhead for encoding candidate articles during a user request is essentially negligible. For scenarios requiring on-the-fly encoding, the complexity for a single news item is governed by the self-attention mechanism at  $\mathcal{O}(L^2D)$ , scaling linearly to  $\mathcal{O}(\mathcal{N}_c \cdot L^2D)$  for a pool of  $\mathcal{N}_c$  candidates. Furthermore, the final ranking process, which relies on vector dot products, requires only  $\mathcal{O}(D)$  per item, resulting in a total scoring complexity of  $\mathcal{O}(\mathcal{N}_c \cdot D)$ .

## 5. Experimental Setup

### 5.1. Dataset Description

We conduct experiments on the public datasets MIND [10] and Adressa [27] to verify the adaptability and robustness of the proposed method in diverse news recommendation scenarios. MIND is a large-scale public news recommendation dataset collected by Microsoft and contains complete user interaction logs from Microsoft News, including impression events and historical click behavior.

- MIND-large: The data spans six weeks (12 October to 22 November 2019) and includes two main components: (1) news content (headlines and summaries) and (2) user interaction records.
- MIND-small is a stratified sample of MIND-large that consists of one week of behavioral data from 50,000 randomly chosen users and maintains consistency with the original data distribution.

The Adressa dataset is a news dataset released in collaboration between the Norwegian University of Science and Technology and the mainstream Norwegian newspaper Adressavisen. The core data include session start and end times, user IDs, news IDs, timestamps, news titles, and news summaries, and cover user behavior over a one-week period (1 January to 7 January 2017). Table 4 provides detailed statistics on key information from the datasets, including the number of users, news articles, and click behaviors.

**Table 4.** Dataset statistics.

| Dataset    | News    | Users     | Interaction Records |
|------------|---------|-----------|---------------------|
| MIND-large | 161,013 | 1,000,000 | 24,155,470          |
| MIND-small | 65,238  | 50,000    | 347,727             |
| Adressa    | 14,732  | 537,627   | 2,527,571           |

## 5.2. Experimental Protocol

Following established practices in news recommendation research [9,28,29], we evaluate all methods on the predefined train/validation/test partitions and keep the validation split exclusively for model selection. In the implementation configuration, MIND-small contains 156,965 training impressions, 73,154 validation impressions, and 73,154 test impressions; MIND-large contains 2,232,748 training impressions, 376,471 validation impressions, and 2,370,727 test impressions; and Adressa contains 145,126 training impressions, 63,226 validation impressions, and 252,902 test impressions. The default reproducibility seed is 3407. For the repeated-run results marked with  $\star$ , we use three independent seeds (3407, 3408, and 3409) and report the mean and standard deviation. The default model configuration uses  $\sigma = 7$ ,  $\xi = 1.5$ , and  $\varsigma = 0.5$  unless otherwise stated in the hyperparameter analysis. User click histories are truncated to the 50 most recent clicked news items, and shorter histories are padded to the same length. News titles are truncated or padded to 30 tokens. Word representations are initialized with 300-dimensional GloVe embeddings, and the model dimensionality follows the configuration used in each reported run. The frequency-domain denoising module preserves the embedding dimensionality because both FFT and inverse FFT are applied only along the sequence dimension. We set the dropout probability to 0.2, use Adam with an initial learning rate of  $2 \times 10^{-4}$ , and apply a 10% linear warm-up schedule. During training, each positive clicked news item is paired with  $K = 4$  negative samples drawn from the unclicked candidates in the same impression. Early stopping is applied on the validation AUC with a patience of five epochs, and the checkpoint with the highest validation AUC is selected for final test evaluation.

## 5.3. Evaluation Metrics

We adopt four core metrics to evaluate recommendation accuracy: area under the ROC curve (AUC), mean reciprocal rank (MRR), and normalized discounted cumulative gain at 5 and 10 (nDCG@5 and nDCG@10). These metrics are widely used in news recommendation [30,31].

Because the improvements on MIND are relatively small, we adopt a stricter statistical protocol for all main comparisons. Each method is evaluated on the same test impressions after validation-based checkpoint selection. For repeated runs, we report the mean and standard deviation across independent seeds, the 95% confidence interval computed from the seed-level scores, and the exact two-sided  $p$ -value obtained by comparing FDDP-RN with the strongest baseline under paired observations. We use a paired  $t$ -test when the seed-level paired differences are approximately normal; otherwise, we use the Wilcoxon signed-rank test. The same protocol is applied to both accuracy metrics and cold-start fairness metrics so that improvements in ranking quality and exposure fairness are assessed under the same statistical standard.

To directly address cold-start fairness and long-tail exposure, we additionally report catalog coverage, cold-start coverage, long-tail exposure rate, exposure Gini, average recommendation popularity, cold-start recall, and group recall gap. Catalog coverage measures the fraction of distinct recommended items over the entire candidate catalog. Cold-start coverage measures the same fraction within  $\mathcal{N}^{\text{cold}}$ . Long-tail exposure rate measures the proportion of top- $k$  slots occupied by cold-start items. Exposure Gini quantifies inequality in item-level exposure counts, with lower values indicating fairer exposure. Average recommendation popularity is the mean training-set popularity ceiling  $\eta^*$  of the recommended items, so a lower value means less popularity concentration. Cold-start recall measures the hit ratio of clicked cold-start items in the top- $k$  list, and group recall gap is the absolute recall difference between the popular and cold-start groups. These metrics complement

AUC, MRR, and nDCG by checking whether accuracy gains are accompanied by fairer exposure of sparse-interaction news.

#### 5.4. Baselines

We evaluate our model against representative news recommendation methods:

1. LSTUR [32] constructs comprehensive user profiles via GRU networks for short-term interest extraction and user ID embeddings for long-term preference modeling, integrated through two distinct fusion methods.
2. NAML [11] adopts an attentive multi-view learning framework that aggregates these diverse features into a unified news representation using a view-level attention mechanism.
3. NWT [33] fuses two types of information through joint training of a topic-aware aggregator and a news encoder emphasizing word correlations, and utilizes enhanced word correlations for topic prediction to mine user topic preferences for generating accurate user representations.
4. MINS [12] detects latent interests of a single news item within a session through a parallel interest network, and learns interest representations from related news of the same interest to mine multi-interest features of the session sequence.
5. PENR [23] incorporates popularity prediction as an auxiliary task by calculating the click counts of each news article to represent its popularity.
6. MCCM [24] generates specific feature extractors for words to capture multi-channel local interest features and incorporates a frequency-aware contrastive learning module to enhance the robustness of user modeling.
7. GLORY [28] constructs a global framework to enhance the local news context, leveraging a globally aware graph neural network and multi-view learning to capture news connections across users.
8. PAD-MPFN [21] mitigates popularity bias by integrating multi-source user interests through sequential behavior, social propagation graphs, and logarithmic time-decay factors and by utilizing dynamic gating mechanisms for personalized weighted mining of user interests.

For baseline implementation sources, LSTUR and NAML use the official MIND implementations; PENR, GLORY, and MCCM use the code released by their authors; and NWT, MINS, and PAD-MPFN are reimplemented following their original papers. All baselines are retrained and reevaluated with the same train/validation/test partitions, title-only input, candidate impressions, negative sampling ratio, evaluation code, early-stopping criterion, and validation-based model-selection rule used for FDDP-RN.

## 6. Results

### 6.1. Overall Performance

Table 5 reports the performance on both MIND datasets. Values are percentages, with the best results bolded and the second-best underlined.

First, models that incorporate auxiliary tasks or multi-source information, such as PENR, NWT, and MINS, generally outperform LSTUR and NAML. PENR leverages a popularity prediction task, NWT enhances representation through topic awareness, and MINS mines multi-interest features via parallel networks. This suggests that single temporal sequence modeling is insufficient to capture the complex evolution of user interests, and multi-dimensional feature enhancement is crucial. Among all baselines, MCCM and PAD-MPFN stand out: PAD-MPFN enhances hidden interests through a global click graph, while MCCM captures multi-granular preferences using a hierarchical interest tree. This validates the effectiveness of structured information and global context for improving recommendation quality.

FDDP-RN achieves the best results overall. The gains mainly come from two synergistic factors. First, the Fourier frequency partitioning mechanism finely filters semantic redundancy in word sequences, enhancing the discriminative power of core keywords. Second, the popularity bias correction mechanism dynamically recalibrates the embedding distribution of niche news, mitigating overfitting to popular signals. Consequently, FDDP-RN not only identifies users' true interests more accurately but also achieves a better balance between recommendation accuracy and fairness.

**Table 5.** Overall performance comparison on the MIND dataset. Best results are bolded, second-best are underlined. \* marks methods with repeated-seed results following the statistical protocol.

| Method     | MIND-Small   |              |              |              | MIND-Large   |              |              |              |
|------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
|            | AUC          | MRR          | nDCG@5       | nDCG@10      | AUC          | MRR          | nDCG@5       | nDCG@10      |
| LSTUR      | 65.87        | 30.78        | 35.15        | 40.15        | 67.08        | 32.36        | 35.15        | 40.93        |
| NAML       | 66.12        | 31.53        | 34.88        | 41.09        | 66.46        | 32.75        | 35.66        | 41.40        |
| NWT *      | 66.19        | 31.70        | 34.97        | 41.22        | 68.50        | 33.55        | 37.17        | 43.41        |
| MINS *     | 67.10        | 31.71        | 35.25        | 41.50        | 68.11        | 32.49        | 34.67        | 42.42        |
| PENR *     | 67.16        | 31.75        | 34.36        | 40.82        | 69.25        | 34.16        | 37.31        | 43.04        |
| MCCM       | 67.95        | 32.76        | 36.62        | 42.66        | <u>69.45</u> | <u>34.41</u> | 37.62        | 43.31        |
| GLORY *    | 67.68        | 32.45        | 35.78        | 42.10        | <u>69.04</u> | 33.83        | 37.53        | 43.69        |
| PAD-MPFN * | <u>68.03</u> | <u>32.90</u> | <u>36.67</u> | <u>42.77</u> | 69.40        | 34.37        | <u>37.66</u> | <u>43.79</u> |
| FDDP-RN *  | <b>68.09</b> | <b>32.96</b> | <b>36.71</b> | <b>42.89</b> | <b>69.53</b> | <b>34.45</b> | <b>37.84</b> | <b>43.86</b> |

Table 6 shows the results on the Adressa dataset. PAD-MPFN again achieves the second-best performance, confirming the effectiveness of global news graph modeling. FDDP-RN achieves further improvements on all metrics, attributed to its accurate suppression of textual noise and distribution calibration of the embedding space. By introducing frequency-domain inductive bias, the model filters semantic redundancy and enhances the discriminative power of abrupt keyword changes. Moreover, given the strong timeliness and cold-start nature of the Adressa dataset, the popularity bias correction factor  $\zeta^{\text{cold}}$  dynamically recalibrates the embedding distribution of cold-start news. This enables the model to accurately capture users' immediate preferences while overcoming the overfitting to popular signals, thereby achieving a better balance between accuracy and novelty. Table 7 then reports the statistical-significance results for all three datasets.

## 6.2. Ablation Study

We conduct component-level and mechanism-level ablations on MIND-small and Adressa. Table 8 reports exact AUC and nDCG@10 values for the complete model, component-removal settings, low-pass, high-pass, band-pass, fixed-scaling, learnable frequency-wise filtering, and simpler PBC alternatives. These results verify that the gains come from the specific frequency-domain filtering and popularity-bias calibration designs rather than from parameter addition alone. We report ablation results on MIND-small and Adressa because they represent two distinct news recommendation settings and allow the component effects to be evaluated across different data distributions. MIND-large is reserved for the overall performance comparison due to its substantially higher computational cost.

- FDDP-RN: Complete FDDP-RN implementation with learnable FDD and norm-rescaling PBC.
- w/o FDD: Remove frequency-domain denoising (disable Algorithm 1 line 9–line 11).
- w/o PBC: Remove popularity bias correction (disable Algorithm 1 line 16).
- w/NE: Remove both frequency-domain denoising and popularity bias correction, using only the base news encoder.

**Table 6.** Overall performance comparison on the Adressa dataset. \* marks methods with repeated-seed statistics.

| Method    | Adressa      |              |              |              |
|-----------|--------------|--------------|--------------|--------------|
|           | AUC          | MRR          | nDCG@5       | nDCG@10      |
| LSTUR     | 73.17        | 39.47        | 39.79        | 47.21        |
| NAML      | 73.14        | 38.10        | 38.56        | 46.13        |
| GLORY     | 74.13        | 42.29        | 43.37        | 49.51        |
| PENR      | 73.84        | 41.87        | 42.56        | 48.16        |
| PAD-MPFN  | <u>74.62</u> | <u>42.90</u> | <u>43.62</u> | <u>49.75</u> |
| FDDP-RN * | <b>75.36</b> | <b>43.17</b> | <b>43.82</b> | <b>50.11</b> |

**Table 7.** Statistical significance test results (FDDP-RN vs. the strongest baseline on each dataset). 95% confidence intervals and paired *p*-values are reported. These statistics are based on repeated runs with different random seeds.

| Dataset    | Metric  | FDDP-RN (Mean $\pm$ std) | Baseline (Mean $\pm$ std) | 95% CI (FDDP-RN) | <i>p</i> -Value |
|------------|---------|--------------------------|---------------------------|------------------|-----------------|
| MIND-small | AUC     | 68.09 $\pm$ 0.04         | 68.03 $\pm$ 0.05          | [68.04, 68.14]   | 0.041           |
|            | MRR     | 32.96 $\pm$ 0.03         | 32.90 $\pm$ 0.04          | [32.92, 33.00]   | 0.033           |
|            | nDCG@5  | 36.71 $\pm$ 0.05         | 36.67 $\pm$ 0.04          | [36.65, 36.77]   | 0.047           |
|            | nDCG@10 | 42.89 $\pm$ 0.05         | 42.77 $\pm$ 0.06          | [42.83, 42.95]   | 0.018           |
| MIND-large | AUC     | 69.53 $\pm$ 0.04         | 69.45 $\pm$ 0.04          | [69.48, 69.58]   | 0.026           |
|            | MRR     | 34.45 $\pm$ 0.03         | 34.41 $\pm$ 0.04          | [34.42, 34.48]   | 0.039           |
|            | nDCG@5  | 37.84 $\pm$ 0.05         | 37.66 $\pm$ 0.04          | [37.78, 37.90]   | 0.022           |
|            | nDCG@10 | 43.86 $\pm$ 0.04         | 43.79 $\pm$ 0.05          | [43.81, 43.91]   | 0.033           |
| Adressa    | AUC     | 75.36 $\pm$ 0.05         | 74.62 $\pm$ 0.06          | [75.30, 75.42]   | 0.004           |
|            | MRR     | 43.17 $\pm$ 0.04         | 42.90 $\pm$ 0.05          | [43.12, 43.22]   | 0.012           |
|            | nDCG@5  | 43.82 $\pm$ 0.05         | 43.62 $\pm$ 0.05          | [43.76, 43.88]   | 0.021           |
|            | nDCG@10 | 50.11 $\pm$ 0.05         | 49.75 $\pm$ 0.06          | [50.05, 50.17]   | 0.009           |

As further shown in Table 8, the learnable FDD design outperforms all simplified frequency variants. Low-pass only, high-pass only, and band-pass only remove complementary information and lead to clear performance drops. Fixed scaling is competitive but remains inferior to the learnable shared scalar. The learnable frequency-wise filter also remains below FDDP-RN on both datasets, showing that the gain does not come from adding a larger set of frequency parameters. For PBC, norm rescaling obtains the best AUC and nDCG@10 on both datasets, outperforming popularity normalization, IPW reweighting, popularity regularization, and the direct removal of PBC.

**Table 8.** Extended ablation results of the FDD and PBC designs on MIND-small and Adressa.

| Variant                         | MIND-Small |         | Adressa |         |
|---------------------------------|------------|---------|---------|---------|
|                                 | AUC        | nDCG@10 | AUC     | nDCG@10 |
| FDD Module                      |            |         |         |         |
| FDDP-RN (learnable)             | 68.09      | 42.89   | 75.36   | 50.11   |
| Low-pass only                   | 66.15      | 40.23   | 73.05   | 47.52   |
| High-pass only                  | 65.87      | 39.85   | 74.42   | 49.02   |
| Band-pass only                  | 67.34      | 41.88   | 74.31   | 48.91   |
| Fixed scaling                   | 67.82      | 42.51   | 74.88   | 49.53   |
| Learnable frequency-wise filter | 67.93      | 42.66   | 75.08   | 49.84   |
| w/o FDD                         | 67.11      | 40.99   | 74.52   | 49.23   |
| w/ NE                           | 65.86      | 40.16   | 74.15   | 48.89   |

Table 8. Cont.

| Variant                | MIND-Small |         | Adressa |         |
|------------------------|------------|---------|---------|---------|
|                        | AUC        | nDCG@10 | AUC     | nDCG@10 |
| PBC Module             |            |         |         |         |
| FDDP-RN (norm rescale) | 68.09      | 42.89   | 75.36   | 50.11   |
| Pop normalization      | 67.55      | 42.15   | 74.58   | 49.08   |
| IPW reweighting        | 67.72      | 42.42   | 74.82   | 49.68   |
| Pop regularization     | 67.68      | 42.38   | 74.15   | 48.85   |
| w/o PBC                | 67.85      | 42.51   | 74.96   | 49.72   |

### 6.3. Hyperparameter Analysis

This section examines the effects of the low-frequency component truncation index  $\sigma$  and the high-frequency component scaling factor  $\xi$  on the FDDP-RN model. The sensitivity analysis is conducted exclusively on the MIND-small validation set, which shares the same data generation mechanism as MIND-large and Adressa, while keeping the grid search computationally tractable. Based on this analysis, we select  $\sigma = 7$  and  $\xi = 1.5$  as the default configuration. We then apply this same setting to all main experiments across MIND-small, MIND-large, and Adressa without additional dataset-specific tuning. This unified configuration ensures fair and consistent comparisons across the three datasets, while substantially reducing the computational overhead that would otherwise be incurred by exhaustive parameter searches on larger datasets. The experimental results demonstrate that this configuration consistently achieves strong performance across all datasets, confirming that the hyperparameters selected on MIND-small generalize effectively. We note that further tuning on individual datasets could yield marginal additional gains, but the current results already validate the effectiveness of FDDP-RN with a single, cost-efficient configuration.

#### 6.3.1. Hyperparameter Analysis of $\sigma$

Figure 5 illustrates the impact of varying the low-frequency component truncation threshold  $\sigma$  on the performance of the FDDP-RN model. Here  $\xi$  is fixed to 1.5, and all values are measured on the MIND-small validation set. The main observations are as follows: when  $\sigma = 7$ , both AUC and nDCG@10 reach their peaks simultaneously. This indicates that this threshold covers the range of amplitudes for burst keywords and strong verbs in the headlines, allowing the FDDP-RN to maximize the retention of core components reflecting news timeliness and click intent, while minimizing redundant information. When  $\sigma = 9$ , the amplitudes of a large number of fluctuating components are truncated. The loss of information from these fluctuating components prevents the model from identifying the core information in the headlines.

#### 6.3.2. Hyperparameter Analysis of $\xi$

To assess the effectiveness of the high-frequency scaling factor  $\xi$ , we conducted hyperparameter experiments using the determined low-frequency cutoff  $\sigma = 7$  and the discrete set of values  $\xi \in \{0.0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0\}$ ; the results are shown in Figure 6. Thus, the conclusion  $\xi = 1.5$  should be understood as the best value in the one-dimensional sweep under  $\sigma = 7$ , not as the global optimum over the full two-dimensional grid. The results reveal a consistent trend: as  $\xi$  increases, model performance first rises sharply to a peak, after which it begins to decline. Specifically, as  $\xi$  increases from 0.0 to 1.5, both AUC and nDCG@10 show a continuous upward trend, reaching the best one-dimensional performance at  $\xi = 1.5$ . However, when  $\xi$  exceeds 1.5, both metrics decline significantly. Specifically, when  $\xi = 0.0$ , the model retains only low-frequency components, resulting in AUC and nDCG@10 reaching their minimum values across the entire range. This confirms that local semantic variations are the core carriers of key semantic differences in

news recommendation, and discarding them severely weakens the model's discriminative ability. When  $\zeta = 1.0$ , the high-frequency components are not scaled, and the original frequency-domain representation is retained. In this case, the model's performance is markedly inferior to the optimal performance achieved at  $\zeta = 1.5$ . This validates that the scaling coefficient  $\zeta$  can effectively enhance the semantic discriminative power of news embeddings by adaptively amplifying high-frequency components under the optimal  $\sigma$  configuration. When  $\zeta > 1.5$ , excessive amplification of high-frequency components disrupts the stability of global semantic modeling and leads to a continuous decline in recommendation performance. Overall, the experiments identify  $\zeta = 1.5$  as the stable default under  $\sigma = 7$ .

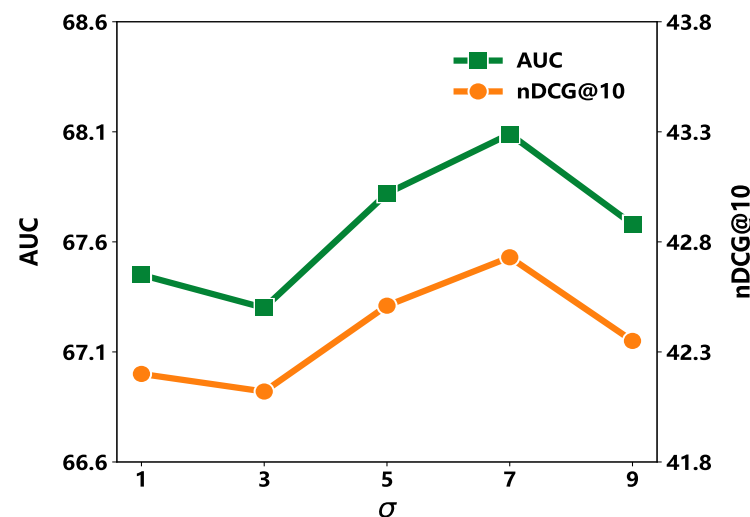


**Figure 5.** The impact of low-frequency component truncation range  $\sigma$  on the MIND-small validation set with  $\zeta = 1.5$ .

### 6.3.3. Hyperparameter Analysis of $\sigma$ and $\zeta$

To evaluate the optimal choices for the parameters  $\sigma$  and  $\zeta$  in FDDP-RN, we conducted experiments using different parameter combinations. Figure 7 presents the experimental results in a visual format. In particular, Figure 7a,b exhibit similar distribution patterns. The two-dimensional grid reaches its highest validation scores at  $\sigma = 5, \zeta = 2$ , with AUC 68.26 and nDCG@10 42.78. The final default setting is nevertheless kept at  $\sigma = 7, \zeta = 1.5$  because it is the simultaneous optimum in the two one-dimensional sweeps and achieves AUC 68.09 and nDCG@10 42.73, only 0.17 AUC and 0.05 nDCG@10 below the joint-grid peak. This default is therefore a conservative stability choice: it avoids selecting a sharper local grid peak while preserving almost the same validation performance. If  $\sigma$  is too small, the model inadvertently incorporates excessive high-frequency information and redundant semantic fluctuations. Furthermore, if  $\sigma$  is too large, it indicates excessive truncation of frequency components, leading to the loss of key semantic information, such as keywords in news headlines, thereby affecting recommendation performance. Furthermore, the high-frequency component scaling factor  $\zeta$  essentially regulates the intensity of local semantic information enhancement. If its value is too small, key discriminative features may be suppressed. Conversely, if its value is too large, high-frequency information may be amplified, thereby disrupting the stability of global semantic information and leading to a decline in performance. Only moderate values of  $\sigma$  and  $\zeta$  can enhance recommendation performance.

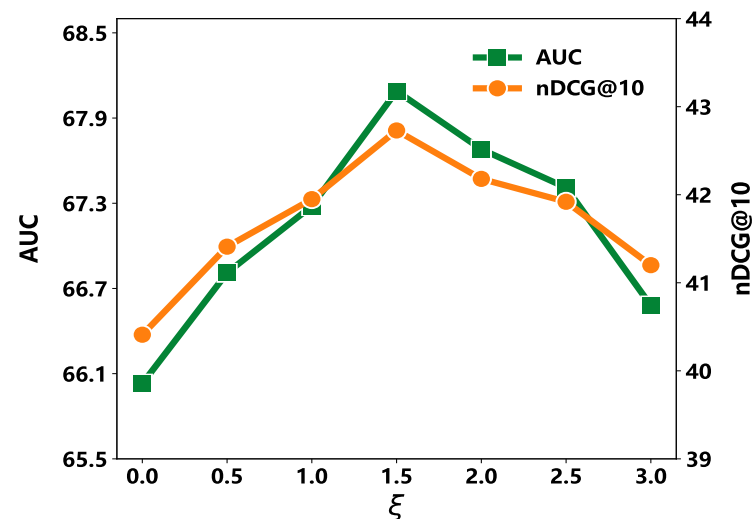


**Figure 6.** The impact of frequency-domain scaling factor range  $\xi$  on the MIND-small validation set with  $\sigma = 7$ .

#### 6.4. Diversity Analysis

This section evaluates recommendation-list diversity on the MIND-small dataset using intra-list minimum distance (ILMD) and intra-list average distance (ILAD). For a top- $k$  recommendation list, ILMD measures the minimum pairwise cosine distance among the normalized embeddings of the recommended news items, whereas ILAD measures their average pairwise cosine distance. Therefore, higher ILMD and ILAD values indicate that the recommended items are less semantically redundant and that the list contains more diverse content. As shown in Table 9, FDDP-RN consistently outperforms the two baseline methods across all evaluation metrics. Specifically, compared with the best-performing baseline PAD-MPFN, FDDP-RN increases ILMD@5 by 0.77%, ILAD@5 by 0.41%, ILMD@10 by 0.86%, and ILAD@10 by 0.50%. These consistent and substantial improvements clearly demonstrate the effectiveness of our FDDP-RN in enhancing the overall diversity of news recommendation lists. Although PAD-MPFN successfully simulates the dynamic evolution of news popularity through first-order system response theory, its understanding of news semantics is still limited to the temporal domain, resulting in insufficient news modeling ability when processing news with rapid word-frequency variations. FDDP-RN improves the matching accuracy between long-tail news and user interests by mining core news semantics through frequency-domain denoising, while simultaneously using bias correction to calibrate the model magnitude of cold-start news. This gives niche cold-start news more exposure opportunities and improves recommendation diversity. As a result, FDDP-RN is particularly effective in enhancing recommendation diversity. This more comprehensive analysis and bias correction strategy enables it to provide richer and more diverse recommendations, resulting in a better user experience.

**Table 9.** Comparison of recommendation diversity. Bold numbers indicate the best results.

| Method   | ILMD@5        | ILAD@5        | ILMD@10       | ILAD@10       |
|----------|---------------|---------------|---------------|---------------|
| PENR     | 20.73%        | 32.17%        | 15.70%        | 31.87%        |
| PAD-MPFN | 21.36%        | 32.84%        | 17.68%        | 32.36%        |
| FDDP-RN  | <b>22.13%</b> | <b>33.25%</b> | <b>18.54%</b> | <b>32.86%</b> |

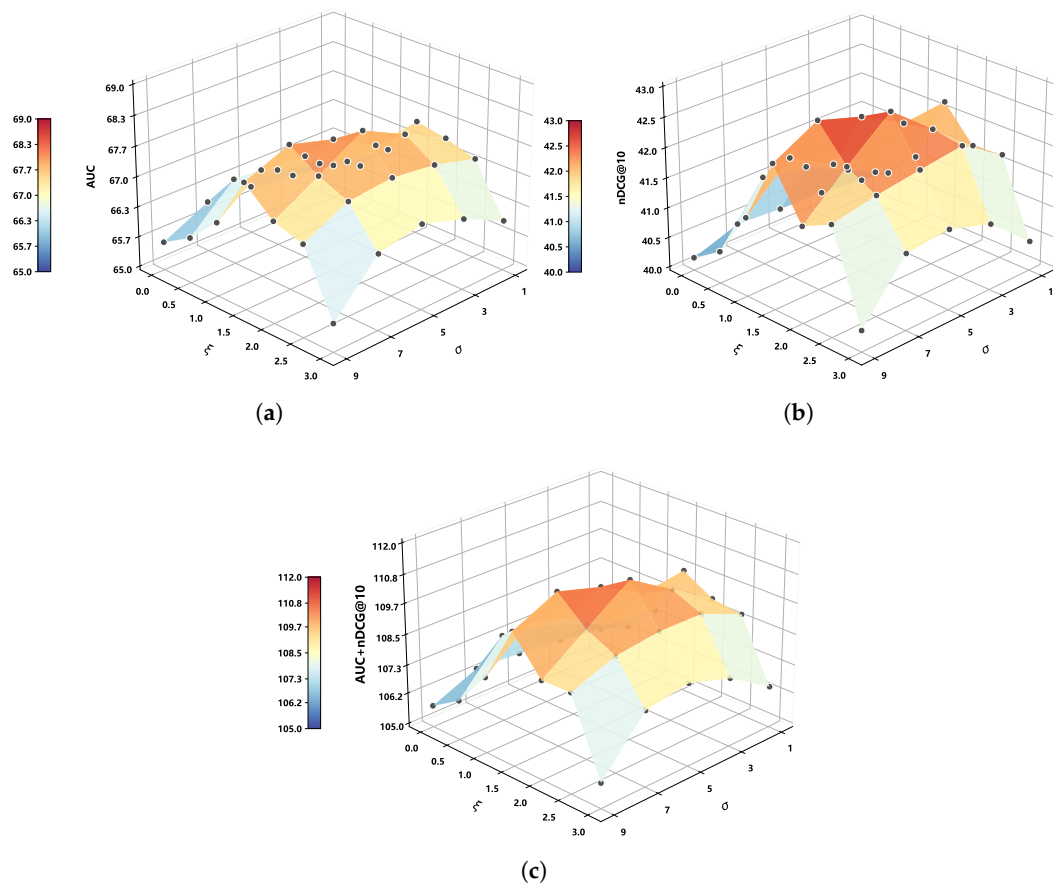


**Figure 7.** Sensitivity analysis of parameters  $\sigma$  and  $\xi$  on the MIND-small validation set: (a) effects of  $\sigma$  and  $\xi$  on FDDP-RN in terms of AUC; (b) effects of  $\sigma$  and  $\xi$  on FDDP-RN in terms of nDCG@10; and (c) validation-based selection of  $\sigma$  and  $\xi$ . Each dot represents one evaluated parameter combination in the grid, and the color scale encodes the corresponding validation score.

### 6.5. Cold-Start Fairness Analysis

To comprehensively assess the effectiveness of the popularity bias correction (PBC) module in mitigating cold-start unfairness, we introduce a suite of fairness-oriented evaluation metrics. Beyond standard accuracy metrics, we evaluate catalog coverage, cold-start coverage, long-tail exposure rate, exposure Gini, average recommendation popularity, and cold-start recall at the top-5 and top-10. These metrics are strictly computed on the test split, utilizing the cold-start partition derived exclusively from training-stage propagation-state clustering to prevent any data leakage.

As presented in Table 10, the complete FDDP-RN model consistently outperforms the ablated variant (w/o PBC) across all fairness indicators on the MIND-small dataset. Specifically, at top-10 recommendations, the introduction of the PBC module increases cold-start coverage from 5.7 to 6.6 and the long-tail exposure rate from 15.3 to 18.1. Furthermore, the exposure Gini coefficient decreases from 0.991 to 0.987, and the average recommendation popularity significantly drops from 88.1 to 79.4. These results clearly demonstrate that the PBC mechanism successfully suppresses the over-recommendation of trending news and significantly boosts the exposure of marginalized cold-start content, thereby achieving a fairer and more diverse recommendation distribution. To further assess the robustness of these fairness improvements, we repeat the fairness evaluation with three independent random seeds (3407, 3408, and 3409). Table 11 reports the mean, standard deviation, 95% confidence interval, and paired two-sided  $p$ -value for the top-10 fairness metrics. The improvements remain statistically significant across all reported fairness indicators. In par-

ticular, the long-tail exposure rate increases from  $15.3 \pm 0.2$  to  $18.1 \pm 0.3$  ( $p = 0.006$ ), while exposure Gini decreases from  $0.991 \pm 0.001$  to  $0.987 \pm 0.001$  ( $p = 0.024$ ). These repeated-run results indicate that the fairness gains are not caused by a single random seed.

**Table 10.** Cold-start fairness results on MIND-small.

| Method  | k  | Catalog Cov. | Cold Cov. | Long-Tail Exp. | Exposure Gini | Avg. Pop. | Cold Recall |
|---------|----|--------------|-----------|----------------|---------------|-----------|-------------|
| w/o PBC | 5  | 16.8         | 3.7       | 12.3           | 0.994         | 130.0     | 23.1        |
| FDDP-RN | 5  | 17.5         | 4.4       | 14.7           | 0.991         | 118.6     | 25.2        |
| w/o PBC | 10 | 26.1         | 5.7       | 15.3           | 0.991         | 88.1      | 47.3        |
| FDDP-RN | 10 | 27.4         | 6.6       | 18.1           | 0.987         | 79.4      | 49.8        |

**Table 11.** Repeated-run statistical analysis of top-10 cold-start fairness metrics on MIND-small. Values are computed over three independent seeds. For Exposure Gini and Avg. Pop., lower values indicate fairer exposure; for the remaining metrics, higher values are better.

| Metric                     | w/o PBC           | FDDP-RN           | 95% CI<br>(FDDP-RN) | p-Value |
|----------------------------|-------------------|-------------------|---------------------|---------|
| Catalog Coverage@10        | $26.1 \pm 0.3$    | $27.4 \pm 0.2$    | [27.15, 27.65]      | 0.018   |
| Cold-start Coverage@10     | $5.7 \pm 0.1$     | $6.6 \pm 0.1$     | [6.48, 6.72]        | 0.011   |
| Long-tail Exposure Rate@10 | $15.3 \pm 0.2$    | $18.1 \pm 0.3$    | [17.73, 18.47]      | 0.006   |
| Exposure Gini@10           | $0.991 \pm 0.001$ | $0.987 \pm 0.001$ | [0.986, 0.988]      | 0.024   |
| Avg. Pop.@10               | $88.1 \pm 1.4$    | $79.4 \pm 1.1$    | [78.03, 80.77]      | 0.009   |
| Cold Recall@10             | $47.3 \pm 0.4$    | $49.8 \pm 0.3$    | [49.43, 50.17]      | 0.013   |

### 6.6. Case Study

In this section, to further verify the effectiveness of popularity bias correction in FDDP-RN, we selected a representative user whose historical click records include two cold-start news articles. Table 12 details the user's historical click sequence and the click counts for the corresponding news articles. The user's click records show that their interests are mainly concentrated in sports (such as N53024, N52897, and the cold-start news N48993). In the uncorrected model, popular news articles often have a larger original vector magnitude in the embedding space due to their large interaction volume. However, the click count of cold-start news articles such as N48993 is only 5. Although it is highly correlated with the user's sports interests, in the traditional model, its contribution to the attention aggregation process is easily ignored because its magnitude is much smaller than that of popular news articles.

Figure 8 illustrates the changes in the distribution of attention weights  $\alpha_{n_i^u}$  for the eight news articles in the user's historical sequence before and after applying the popularity bias correction factor  $\zeta^{\text{cold}}$ . Specifically, although the frequency-domain encoder of FDDP-RN can identify the semantic relevance between popular sports news N53024 and N52897 and assign them higher weights, irrelevant popular news N41089 still occupies a large share of attention due to its large magnitude, constituting significant popularity noise. Meanwhile, the cold-start news N48993, which truly aligns with the user's core preferences, is marginalized due to its extremely small initial magnitude, and its semantic value is severely obscured. After introducing the popularity bias correction factor  $\zeta^{\text{cold}}$ , the weight distribution undergoes a highly semantically oriented reconstruction. The weight of cold-start news N48993 increases significantly, reaching the same order of magnitude as that of popular sports news N52897. This indicates that magnitude calibration successfully eliminates unfair competition between hot and cold samples, enabling the model to allocate attention fairly based on semantic matching. With the restoration of semantic signals, the inflated weight of irrelevant trending news item N41089 is effectively reduced. Notably,

the corrected distribution does not result in a filter bubble of a single category. The weights of secondary interest categories (such as entertainment N52116 and music N36590), which were previously suppressed by trending topics (like N41089), show a slight rebound. In conclusion, FDDP-RN, through the synergistic effect of frequency-domain denoising and magnitude calibration, not only accurately identifies users' core interests but also effectively preserves the ability to perceive long-tail and secondary interests, achieving an excellent balance between accuracy, fairness, and diversity in recommendations.

**Table 12.** Historically clicked news of the case-study user. All eight rows are clicked items from the user's training history. (○: cold-start news; △: popular news).

| News ID | Title                                                                                            | Category  | $\eta$ | Cold-Start News? |
|---------|--------------------------------------------------------------------------------------------------|-----------|--------|------------------|
| N53024  | AP Top 25: LSU jumps to No. 2, upset drops Georgia to No. 10                                     | Sports    | 258    | △                |
| N48993  | Watch: FOX butchers Gerardo Parra 'Baby Shark' song                                              | Sports    | 5      | ○                |
| N7448   | Every 2019 Mid-Size Pickup Truck Ranked                                                          | Autos     | 9      | ○                |
| N52897  | Jeremy Pruitt has hilarious quote about Tennessee's chances of beating Alabama                   | Sports    | 260    | △                |
| N41089  | This Man's Rattlesnake Bite Is a Warning to Everyone to Take Animal Bites More Seriously         | Lifestyle | 1163   | △                |
| N3863   | These 4 words are the key to becoming rich                                                       | News      | 752    | △                |
| N52116  | Bachelor Colton Underwood Calls Out Bachelorette Rachel Lindsay's 'Hypocrisy over the Last Year' | TV        | 523    | △                |
| N36590  | Tim McGraw Proudly Shows Off His Ripped Physique: 'Best Shape of My Life'                        | Music     | 185    | △                |

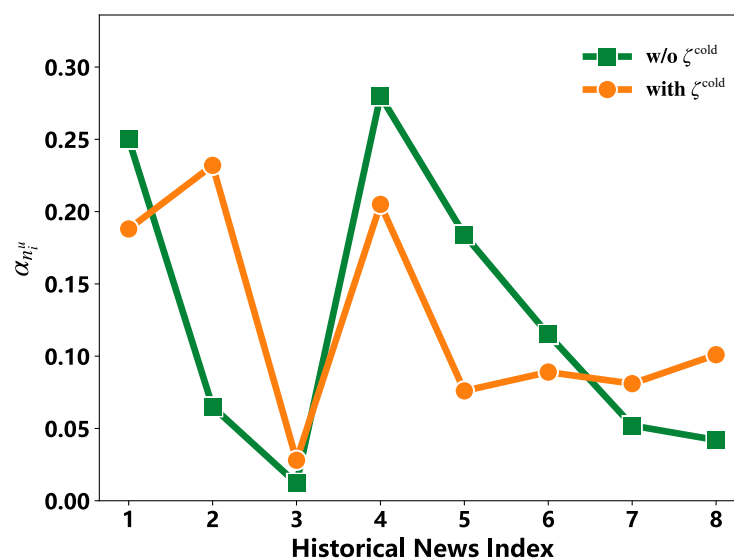


**Figure 8.** Comparison of user attention weights before and after applying popularity bias correction.

## 7. Conclusions and Future Work

In this paper, we propose FDDP-RN, a personalized news recommendation model that combines frequency-domain denoising with embedding-magnitude calibration to address semantic redundancy and cold-start unfairness. First, FFT-based frequency-domain denoising is introduced into the news encoder to preserve global semantic context while enhancing abrupt, high-frequency signals in news titles. Second, the PBC module recalibrates the magnitudes of historical cold-start news embeddings toward the norm benchmark for popular news without changing their semantic directions. Experiments on MIND-small, MIND-large, and Adressa show that FDDP-RN improves ranking accuracy, recommendation diversity, and cold-start exposure.

Despite these advances, we identify several promising directions for future research:

- Deeper use of pre-trained models: to address the limited semantic depth of static word vectors, future work will introduce pre-trained language models as the foundation of the encoder, leverage their contextual modeling capabilities to improve feature extraction accuracy, and explore efficient parameter fine-tuning techniques to balance real-time system performance.
- Information diffusion modeling: future research could further integrate news dissemination patterns and event-evolution trajectories within social networks, particularly for breaking news scenarios, and explore how to enhance the timeliness and foresight of recommendations by modeling the spatiotemporal diffusion patterns of information flow.
- Multimodal and knowledge extension: future work will attempt to integrate multimodal features such as news videos and images, enrich content representation with large-scale knowledge graphs, and explore the zero-shot reasoning capabilities of large language models to provide deeper semantic enhancement for news items in extreme cold-start states.

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

**Funding:** This work was supported in part by the Natural Science Foundation of Fujian Province, China (Grant No. 2022J05176), the President’s Fund of Minnan Normal University (No. KJ2022002), and the Minnan Normal University Advanced Cultivation Project (No. MSGJB2023019).

**Data Availability Statement:** The datasets analyzed in this study are publicly available. The Microsoft News Dataset (MIND) is available at <https://msnews.github.io/>; the Adressa dataset is available from its original public release.

**Conflicts of Interest:** Author Huifan Gao was employed by Xiamen Airlines Co., Ltd. (Xiamen, China). The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest. All research was conducted at Minnan Normal University under the direction of Dr. Biyang Ma.

## References

1. Zhu, Q.; Zhou, X.; Song, Z.; Tan, J.; Guo, L. Dan: Deep attention neural network for news recommendation. In Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence (AAAI), Honolulu, HI, USA, 27 January–1 February 2019; pp. 5973–5980. [CrossRef]

2. Qi, T.; Wu, F.; Wu, C.; Huang, Y. News Recommendation with Candidate-aware User Modeling. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), Madrid, Spain, 11–15 July 2022; pp. 1917–1921. [\[CrossRef\]](#)
3. Kim, T.; Kim, Y.; Lee, Y.C.; Shin, W.Y.; Kim, S.W. Is It Enough Just Looking at the Title?: Leveraging Body Text To Enrich Title Words Towards Accurate News Recommendation. In Proceedings of the 31st ACM International Conference on Information & Knowledge Management (CIKM), Atlanta, GA, USA, 17–21 October 2022; pp. 4138–4142. [\[CrossRef\]](#)
4. Jiang, S.; Song, H.; Lu, Y.; Zhang, Z. News Recommendation Method Based on Candidate-Aware Long- and Short-Term Preference Modeling. *Appl. Sci.* **2025**, *15*, 300. [\[CrossRef\]](#)
5. Li, J.; Zhu, J.; Bi, Q.; Cai, G.; Shang, L.; Dong, Z.; Jiang, X.; Liu, Q. MINER: Multi-Interest Matching Network for News Recommendation. In *Proceedings of the Findings of the Association for Computational Linguistics: ACL 2022*; Association for Computational Linguistics: Dublin, Ireland, 2022; pp. 343–352. [\[CrossRef\]](#)
6. Treviso, M.; Lee, J.U.; Ji, T.; Van Aken, B.; Cao, Q.; Ciosici, M.R.; Hassid, M.; Heafield, K.; Hooker, S.; Raffel, C.; et al. Efficient methods for natural language processing: A survey. *Trans. Assoc. Comput. Linguist.* **2023**, *11*, 826–860. [\[CrossRef\]](#)
7. Lee-Thorp, J.; Ainslie, J.; Eckstein, I.; Ontanon, S. FNet: Mixing Tokens with Fourier Transforms. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*; Association for Computational Linguistics: Seattle, WA, USA, 2022; pp. 4296–4313. [\[CrossRef\]](#)
8. Qi, T.; Wu, F.; Wu, C.; Yang, P.; Yu, Y.; Xie, X.; Huang, Y. 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 (ACL-IJCNLP)*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2021; pp. 5446–5456. [\[CrossRef\]](#)
9. Ding, Y.; Wang, B.; Cui, X.; Xu, M. Popularity prediction with semantic retrieval for news recommendation. *Expert Syst. Appl.* **2024**, *247*, 123308. [\[CrossRef\]](#)
10. Wu, F.; Qiao, Y.; Chen, J.H.; Wu, C.; Qi, T.; Lian, J.; Liu, D.; Xie, X.; Gao, J.; Wu, W.; et al. MIND: A Large-scale Dataset for News Recommendation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 3597–3606. [\[CrossRef\]](#)
11. Wu, C.; Wu, F.; An, M.; Huang, J.; Huang, Y.; Xie, X. Neural News Recommendation with Attentive Multi-View Learning. In Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI), Macao, China, 10–16 August 2019; pp. 3863–3869. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Wang, R.; Wang, S.; Lu, W.; Peng, X. News Recommendation Via Multi-Interest News Sequence Modelling. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, 23–27 May 2022; pp. 7942–7946. [\[CrossRef\]](#)
13. Wang, R.; Wang, S.; Lu, W.; Peng, X.; Zhang, W.; Zheng, C.; Qiao, X. Intention-Aware User Modeling for Personalized News Recommendation. In *Proceedings of the International Conference on Database Systems for Advanced Applications (DASFAA)*; Springer: Cham, Switzerland, 2023; pp. 373–389. [\[CrossRef\]](#)
14. Rao, Y.; Zhao, W.; Zhu, Z.; Lu, J.; Zhou, J. Global Filter Networks for image classification. *Adv. Neural Inf. Process. Syst. (NeurIPS)* **2021**, *34*, 980–993.
15. Zheng, L.; Lu, C.T.; Jiang, F.; Zhang, J.; Yu, P. Spectral Collaborative Filtering. In Proceedings of the 12th ACM Conference on Recommender Systems (RecSys), Vancouver, BC, Canada, 2 October 2018; pp. 311–319. [\[CrossRef\]](#)
16. Wei, L.; Ni, R.; Wei, J.; Jiang, Y. Frequency-sensitive diffusion model for personalized sequential recommendation. *Neurocomputing* **2025**, *654*, 131313. [\[CrossRef\]](#)
17. Yang, W.; Zhong, R.; Chen, Y.; Li, S.; Ping, H.; Lu, C.; Jiang, P. FITMM: Adaptive Frequency-Aware Multimodal Recommendation via Information-Theoretic Representation Learning. In Proceedings of the 33rd ACM International Conference on Multimedia (MM), Dublin, Ireland, 27–31 October 2025. [\[CrossRef\]](#)
18. Zhao, Q.; Chen, X.; Zhang, H.; Li, X. Dynamic Hierarchical Attention Network for news recommendation. *Expert Syst. Appl.* **2024**, *255*, 124667. [\[CrossRef\]](#)
19. Qi, T.; Wu, F.; Wu, C.; Huang, Y. PP-Rec: News Recommendation with Personalized User Interest and Time-aware News Popularity. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP), Online, 1–6 August 2021; pp. 5457–5467. [\[CrossRef\]](#)
20. Meng, L.; Shi, C.; Hao, S.; Su, X. DCAN: Deep Co-Attention Network by Modeling User Preference and News Lifecycle for News Recommendation. In *Proceedings of the International Conference on Database Systems for Advanced Applications (DASFAA)*; Springer: Cham, Switzerland, 2021; pp. 681–696. [\[CrossRef\]](#)
21. Ma, B.; Deng, Y.; Gao, H. PAD-MPFN: Dynamic Fusion with Popularity Decay for News Recommendation. *Electronics* **2025**, *14*, 3057. [\[CrossRef\]](#)
22. Yang, J. Effects of Popularity-Based News Recommendations (“Most-Viewed”) on Users’ Exposure to Online News. *Media Psychol.* **2016**, *19*, 243–271. [\[CrossRef\]](#)

23. Wang, J.; Chen, Y.; Wang, Z.; Zhao, W. Popularity-enhanced news recommendation with multi-view interest representation. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management (CIKM), Virtual, 1–5 November 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 1949–1958. [[CrossRef](#)]
24. Wang, J.; Jiang, Y.; Li, H.; Zhao, W. Improving news recommendation with channel-wise dynamic representations and contrastive user modeling. In Proceedings of the 16th ACM International Conference on Web Search and Data Mining (WSDM), Singapore, 27 February–3 March 2023; pp. 562–570. [[CrossRef](#)]
25. Pennington, J.; Socher, R.; Manning, C.D. GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*; Association for Computational Linguistics: Doha, Qatar, 2014; pp. 1532–1543. [[CrossRef](#)]
26. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, 4–9 December 2017; Volume 30.
27. Gulla, J.A.; Zhang, L.; Liu, P.; Özgübek, Ö.; Su, X. The Adressa Dataset for News Recommendation. In Proceedings of the International Conference on Web Intelligence (WI), Leipzig, Germany, 23–26 August 2017; pp. 1042–1048. [[CrossRef](#)]
28. Yang, B.; Liu, D.; Suzumura, T.; Dong, R.; Li, I. Going Beyond Local: Global Graph-Enhanced Personalized News Recommendations. In *Proceedings of the 17th ACM Conference on Recommender Systems (RecSys)*, Singapore, 18–22 September 2023; Association for Computing Machinery: New York, NY, USA, 2023; pp. 24–34. [[CrossRef](#)]
29. Yang, Z.; Wang, W.; Qi, T.; Zhang, P.; Zhang, T.; Zhang, R.; Liu, J.; Huang, Y. GLoCIM: Global-view Long Chain Interest Modeling for news recommendation. In *Proceedings of the 2024 Conference on Computational Linguistics (COLING)*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2024; pp. 5084–5095.
30. Wang, S.; Guo, S.; Wang, L.; Liu, T.; Xu, H. HDNR: A hyperbolic-based debiased approach for personalized news recommendation. In Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), Taipei, Taiwan, 23–27 July 2023; pp. 259–268. [[CrossRef](#)]
31. Qi, T.; Wu, F.; Wu, C.; Huang, Y.; Xie, X. Privacy-Preserving News Recommendation Model Learning. In *Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2020*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 1423–1432. [[CrossRef](#)]
32. An, M.; Wu, F.; Wu, C.; Zhang, K.; Liu, Z.; Xie, X. Neural news recommendation with long-and short-term user representations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 336–345. [[CrossRef](#)]
33. Pu, X.; Zhang, J.; Chen, X.; Qian, Y.; Zhang, R. News Recommendation with Word-Related Joint Topic Prediction. *IEEE Access* **2024**, *12*, 72566–72577. [[CrossRef](#)]

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