Ingénierie des Systèmes d'Information
Vol. 31, No. 6, June, 2026, pp. 2035-2047
Journal homepage: http://iieta.org/journals/isi

PA-TimeMF: A Unified Bias-Aware Temporal Matrix Factorization Framework for Dynamic News Recommendation

Z. K. A. Baizal*, Nadia Astria Savitri

School of Computing, Telkom University, Bandung 40257, Indonesia

Corresponding Author Email: baizal@telkomuniversity.ac.id

https://doi.org/10.18280/isi.310624

Received: 29 January 2026
Revised: 30 March 2026
Accepted: 13 April 2026
Available online: 30 June 2026

Keywords:
news recommendation system, bias-aware recommendation, temporal matrix factorization, position bias correction, implicit feedback learning, intelligent information systems, temporal dynamics modeling

ABSTRACT

News recommendation systems play an important role in helping users efficiently access relevant information from rapidly growing digital content. However, implicit feedback-based recommendation models often suffer from biased click signals caused by item position effects and rapidly changing user interests due to the short lifecycle of news articles. Existing approaches generally address position bias correction and temporal dynamics separately, which may lead to inconsistent learning signals. This study proposes position-aware temporal matrix factorization (PA-TimeMF), a unified bias-aware temporal matrix factorization framework for dynamic news recommendation. The proposed model integrates inverse propensity scoring for correcting position bias in clicked interactions and temporal weighting for distinguishing meaningful negative feedback from missing interactions in non-clicked samples. Experiments conducted on the EB-NeRD dataset demonstrate that PA-TimeMF consistently outperforms conventional matrix factorization models, individual debiasing methods, and existing temporal recommendation approaches. The proposed framework achieves improvements of up to 57.7% in HR@10 compared with the baseline model, while maintaining the computational efficiency and interpretability of matrix factorization. The findings indicate that jointly modeling bias correction and temporal evolution provides a more reliable solution for recommendation systems operating in dynamic information environments. This research offers a practical approach for developing accurate and adaptive intelligent information systems for large-scale news platforms.

1. INTRODUCTION

Advances in information technology have driven a rapid increase in the production and consumption of digital content. This growth has also occurred in the news domain. Digital news platforms produce thousands of articles every day and create a variety of highly dynamic and complex information. [1, 2]. A recent bibliometric analysis of articles on artificial intelligence and journalism found a rapid increase. In 2019, the number of articles was 41, and in 2023, the number increased to 122. This data indicates the accelerating transformation of digital articles in the news industry [2]. This situation poses a serious challenge in the form of information overload, where users find it difficult to access relevant information amid abundant available information. Recent studies indicate that 39% of internet users are actively or occasionally avoiding news consumption because of feeling overwhelmed by the amount of information available [3]. This situation highlights the significance of news recommendation systems as the primary mechanism for selecting information. Designing such systems that are both reliable and computationally efficient represents a fundamental challenge in information systems engineering.

However, in the case of news recommendation, the issues are much more complex due to the dynamic nature of news articles. Typically, news articles have a short lifecycle and lose relevance over time, usually within a few hours or days of publication [4]. In addition, the behavior of the user is affected by position bias, which makes the user more likely to click on the items located at the top of the list of recommendations, irrespective of the content's actual relevance [5, 6]. These two factors are fundamental challenges that need to be addressed in an integrated manner.

The problem of position bias has been widely discussed with strong empirical evidence. Collins et al. [5] showed that, for digital library recommendation systems, click-through rates (CTR) for the top-ranked items were 53% higher on Sowiport and 87% higher on JabRef than expected without position bias. Wang et al. [6] proposed a causal inference-based method to distinguish the probability that users examine an item based on its position from the relevance probability that reflects the alignment between the item and the user's needs. By using the Expectation-Maximization (EM) algorithm, the examination probabilities for each position were estimated. After that, the Inverse Propensity Score (IPS) weighting was used to correct the overestimation of the items at the high positions [6].

The temporal dynamics also pose many challenges. Li et al. [7] proposed a dynamic popularity weighting using the TimeMF model, which categorizes articles as HOT and COLD based on their lifecycle stages. The study revealed that the popularity of articles peaks at day 1000. An et al. [4] used the MIND dataset and proved that 70% of clicks happen within the first 24 hours, and 85% of clicks happen within the first 48 hours of publication of articles. These findings confirm that news articles possess extremely short relevance periods, distinct from the e-commerce domain.

Although significant progress has been made in addressing position bias and temporal dynamics individually, most existing methods still treat these two phenomena as separate problems. Chen et al. [8] in a comprehensive survey on debiasing in recommendation systems, noted that the majority of techniques focus on a single type of bias and rarely consider interactions among various bias. Several studies have attempted to overcome this limitation. Zheng et al. [9] proposed a causal inference-based framework to separate user interests and popularity bias, but this required complex neural network architectures and did not explicitly model temporal dynamics. Yang et al. [10] developed a unified framework for multi-bias mitigation, but the evaluation was conducted on e-commerce datasets with significantly longer lifecycles.

The research gap becomes increasingly critical because position bias and temporal dynamics in news recommendation are not mutually independent phenomena [8, 11]. New articles at high positions gain greater exposure, creating a confounding effect that obscures genuine temporal relevance [11, 12]. On the other hand, users show a higher level of exploration when they are consuming very current news [4]. This two-way relationship indicates that modeling these phenomena separately leads to error propagation in the recommendation process [8]. Considering the research gap, this research proposes a framework called position-aware temporal matrix factorization (PA-TimeMF). PA-TimeMF integrates the position bias correction and temporal dynamics modeling into a single objective function. For clicked items, IPS weighting based on EM-estimated examination probabilities corrects overestimation caused by positional advantage [6]. For non-clicked items, dynamic temporal weighting with a 48-hour threshold distinguishes true negatives from missing data [7]. This separation allows positive and negative interaction signals to be corrected independently, without sacrificing the computational efficiency of matrix factorization.

The contributions of this study are as follows: (1) proposing a unified matrix factorization-based framework that distinguishes between click and non-click interactions through separate weighting schemes with IPS for position bias and temporal weighting for missing data; (2) integrating position bias correction and temporal dynamics modeling into a single objective function for simultaneous modeling that avoids error accumulation; (3) comprehensively evaluating the model on the EB-NeRD dataset through ablation studies and comparisons with position bias correction methods, temporal modeling approaches, and state-of-the-art news recommendation systems. By addressing bias and temporal dynamics within a unified and lightweight framework, this research contributes to the broader goal of building reliable information systems capable of delivering relevant content efficiently in dynamic and time-sensitive environments.

2. RELATED WORK

2.1 Collaborative filtering with implicit feedback

Collaborative filtering with implicit feedback has become the dominant approach in recommendation systems because user interaction data can be easily collected without requiring explicit ratings. Hu et al. [13] proposed a matrix factorization framework that treats unobserved entries as negative signals with varying confidence weights. This approach is based on the assumption that the absence of interaction does not always carry the same level of informativeness. Subsequently, Chen et al. [14] introduced FAWMF, which applies adaptive popularity-based weighting, assigning higher weights to popular non-clicked items under the assumption that they were exposed but irrelevant to the user.

Comprehensive comparative studies demonstrate that, with appropriate hyperparameter tuning, matrix factorization approaches can achieve competitive performance even on large-scale datasets [15]. This finding motivates their use as an efficient base model. However, most of these approaches still employ popularity as a proxy for exposure without explicitly modeling position effects. Furthermore, temporal dynamics have not been extensively considered, particularly in domains with rapid content turnover such as news recommendation.

2.2 Position bias in recommendation systems

Position bias has been recognized as a source of systematic distortion in learning from user interactions. Joachims et al. [16] developed propensity-based estimators using IPS to correct position bias without randomized intervention. A method for estimating position bias in personalized search using the EM algorithm was developed by Wang et al. [6], separating examination probability that depends on position from relevance probability that depends on content.

Empirical evidence regarding the magnitude of position bias comes from a study on digital library recommendation systems with over 10 million recommendations. The results showed that CTR for top-ranked items were 53% and 87% higher, respectively, compared to unbiased scenarios [5]. Agarwal et al. [17] proposed estimating position bias without intrusive interventions, using natural query variations instead, making it more feasible for production systems. A tutorial on "Unbiased Learning to Rank" was given by Oosterhuis and de Rijke, in which the debiasing of learning to rank was classified into three types: intervention-based, propensity-based, and dual learning [18].

However, recent research has found that position bias is an issue even in large language model (LLM)-based systems. In their study, Bito et al. [19] found that position bias in LLM-based recommendation systems can reach 25%. This result means that items placed in initial positions have higher selection probabilities regardless of their relevance. Even though position bias research has made tremendous progress in web search and e-commerce, its application, especially with dynamic temporal factors in news recommendation, is still limited.

2.3 Temporal dynamics in recommendation

Temporal dynamics have emerged as an essential factor that distinguishes the news recommendation domain from other domains. The model of recommendation using implicit feedback and time information, named TimeMF, was proposed by Li et al. [7]. It used dynamic popularity weighting based on item classification in the HOT and COLD stages, with a threshold of 1000 days, simulating the long lifecycle of products in e-commerce. The key factors of temporal dynamics in news recommendation, such as content freshness, event evolution, and user interest drift, have been identified in an extensive survey. The authors also emphasized that news articles have extremely short relevance windows and are fundamentally different from e-commerce products [20]. Moreover, empirical evidence shows that 70% of clicks occur within the first 24 hours and 85% within the first 48 hours of publication [4], thus verifying rapid temporal decay as a characteristic of the news domain.

In 2020, Li et al. [21] introduced a time-aware attention mechanism for integrating temporal information into recommendations. However, the process is based on the implicit learning of the temporal weights. Ryu et al. [22] proposed Lifetime-aware Interest Matching (LIME), a lifetime-aware framework for news recommendation incorporating freshness-guided interest refinement, but without addressing position bias. Recent research on the RecSys Challenge 2024, with over 380 million impression logs, has shown that timeliness features play an essential role in the performance of the news recommendation system [23]. This research further supports the idea of explicitly modeling temporal dynamics to distinguish fresh articles from decayed ones.

The one crucial gap that remains unfilled is the lack of consideration of the interplay between temporal dynamics and position bias. The newly posted articles at low positions suffer from a dual disadvantage, while the older articles at high positions continue to receive clicks despite their decreased relevance. The above issue has not been explicitly addressed in any previous temporal models

2.4 News recommendation systems

Recommendation systems for news have moved from simple content-based filtering to sophisticated methods that combine multiple signals. Neural news recommendation with multi-head self-attention, a state-of-the-art model on the MIND dataset, was proposed by Wu et al. [15]. Candidate-aware user modeling, which models both the candidate news and the reading history, was proposed by Qi et al. [24]. DeepFM, which combines factorization machines with deep neural networks, was proposed by Guo et al. [25]. LightGCN showed that simple models can perform better while incurring lower computational costs [26].

Despite the good performance of news recommendation system methods, several methods have limitations, particularly in the news domain. First, the methods are computationally expensive, having millions of parameters, which demand high GPU resources [27]. Second, the methods demand a lot of data, which impose constraints on platforms where data is limited. Third, the methods do not handle bias well, as the click data is assumed to be representative of the true preference without considering position bias [8]. Fourth, the methods are not interpretable, which makes it difficult to understand the reasoning behind the recommendations [15]. As indicated by comparative studies, matrix factorization with proper tuning has the potential to achieve comparable performance with much less computational cost, which forms the basis for the proposed bias correction method [28].

2.5 Multi-bias debiasing approaches

Recently, various studies have acknowledged that the performance of a recommendation system is affected by multiple interacting biases. A comprehensive survey of debiasing in recommendation systems was also presented by Chen et al. [8]. The survey was divided into pre-processing, in-processing, and post-processing methods. The survey also acknowledged that the majority of methods address only a single type of bias and do not account for the complex interactions among different biases. Liu et al. [29] proposed a causal inference framework to disentangle multiple biases in news recommendation. Although the proposed approach has addressed position bias, popularity bias, and freshness bias. It assumes biases are independent and missing the confounding interaction between position bias and temporal dynamics that PA-TimeMF explicitly models. Yang et al. [10] developed a unified multi-bias mitigation framework, but applied a single combined weight to all interactions without distinguishing clicked from non-clicked items, and evaluated only on e-commerce datasets. Chen et al. [30] proposed Time and Content-aware Causal Model (TCCM), which addresses popularity bias by incorporating timeliness into causal inference for news recommendation, but without addressing position bias. These existing approaches demonstrate that multi-bias integration remains an open challenge, particularly in the news domain where position bias and temporal dynamics interact.

In their comprehensive survey on fairness in recommendation systems, Deldjoo et al. highlighted the significance of different types of bias, not only affecting the accuracy of the system but also affecting fairness towards a wide range of stakeholders [31]. Consistent with this, Ekstrand et al. emphasized the importance of responsible recommendation system approaches that can simultaneously address a range of biases while accounting for their ethical implications [27].

Nevertheless, in practice, most approaches to handling multiple biases are based on sequential or additive strategies, which may also bring about compounding errors. Recent approaches such as TCCM [30] address timeliness and popularity bias in news recommendation but leave position bias unmodeled, while LIME [22] focuses on lifecycle-aware temporal dynamics without bias correction. Unlike these approaches, PA-TimeMF simultaneously addresses both through differential weighting preventing compound errors from conflicting correction signals. This also indicates the need for a unified modeling framework, which takes position bias and temporal dynamics simultaneously. For this purpose, PA-TimeMF proposes three significant mechanisms: simultaneous modeling in a single objective function, differential treatment applying IPS weighting only to clicked items and temporal weighting only to non-clicked items, and explicit parameterization for interpretability and computational efficiency.

3. METHODOLOGY

This section introduces a unified framework, PA-TimeMF, that models position bias and temporal dynamics in news recommendation systems. PA-TimeMF integrates two different methods for bias correction. The first method is position bias correction via the EM method proposed by Wang et al. [6]. The second method is modeling temporal dynamics via the TimeMF method proposed by Li et al. [7].

PA-TimeMF is designed to address two key behaviors: users clicking top-positioned articles regardless of relevance, and news articles losing relevance rapidly due to their short lifecycles. News articles experience rapid popularity decay within 48 hours of publication, transitioning from the HOT to the COLD phase. Hence, position bias correction and temporal dynamics modeling must be done simultaneously using a unified learning framework, as shown in Figure 1.

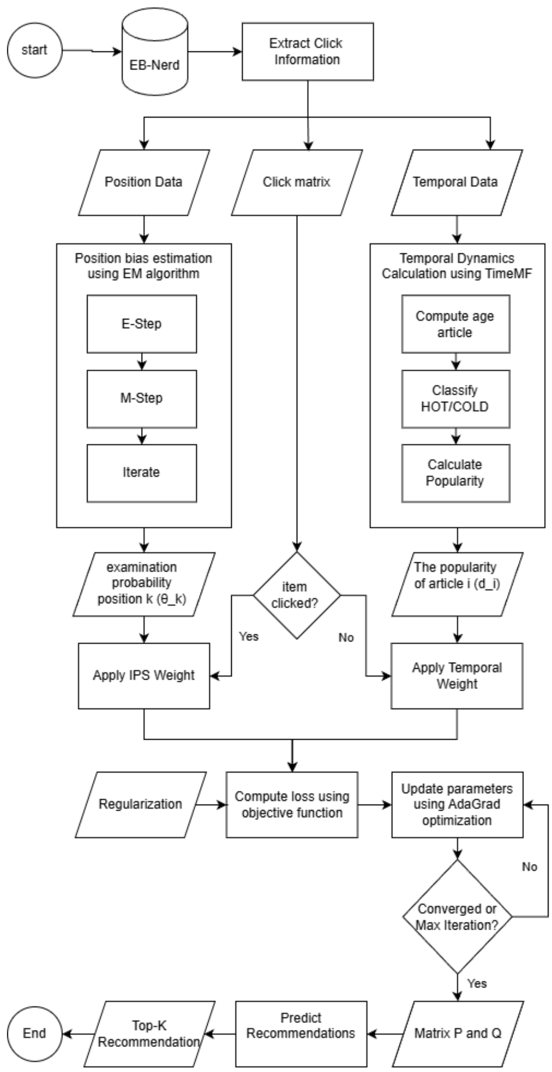


Figure 1. System design of position-aware temporal matrix factorization (PA-TimeMF)

Figure 1 shows the overall system design of the PA-TimeMF system, which has three major parts: data preprocessing, training, and recommendation inference. The EB-NeRD dataset is a source of two major types of data: click data, which represents user-item interactions, and temporal data, which represents article publication time information. Both types of data are handled by two parallel workflows: the position bias correction workflow for computing the probabilities of examination at each position, and the temporal dynamics modeling workflow for computing article popularity weights based on content age. Subsequently, these two pipelines are integrated into a single objective function to produce a recommendation model capable of leveraging click and non-click signals more accurately.

3.1 Problem formulation and notation

Suppose there are M users and N news articles in the system. The user-item interaction matrix $R \in R^{M \times N}$ represents implicit feedback, where input $r_{ui} = 1$ if user u clicks on article i, and $r_{ui} = 0$ if there is no click. Unlike explicit feedback (e.g., ratings), implicit feedback is one-class, meaning it only observes positive signals (clicks) without explicit negative signals [13]. Missing input ($r_{ui} = 0$) are ambiguous and can represent either true negatives (the article is indeed not relevant) or missing data (the article is relevant but was not exposed to the user). Each interaction is associated with a position $p_{ui} \in \{1,2,\ldots,K\}$, which indicates the position of article i when displayed to user u in the recommendation list, as well as a timestamp t_{ui} indicating when the interaction occurred. Each article has a publication time pub_i indicating when the article was published.

The objective of PA-TimeMF is to learn latent factor matrices $P \in \mathbb{R}^{d \times M}$ and $Q \in \mathbb{R}^{d \times N}$, where $p_u \in \mathbb{R}^d$ represents the latent vector of user u, and $q_i \in \mathbb{R}^d$ represents the latent vector of article i. The preference score is predicted as below,

$$\hat{r}_{ui} = \boldsymbol{p}_u^T \boldsymbol{q}_i = \sum_{f=1}^{d} p_{uf} \cdot q_{if} \tag{1}$$

where, d is the dimensionality of the latent space. This model simultaneously corrects position bias (items at higher positions tend to be over-represented in clicks) and models temporal dynamics (newer articles are more relevant than older articles).

3.2 Temporal information analysis

News articles possess temporal characteristics that are fundamentally different from e-commerce products. As reported by An et al. [4], approximately 70% of clicks occur within the first 24 hours and 85% within the first 48 hours following publication, indicating extremely rapid decay in news relevance. This is in sharp contrast to the e-commerce domain, where the products' relevance is maintained for months or even years.

Following Li et al. [7], this research employs a two-stage temporal model: HOT (newly published, high relevance) and COLD (older articles, decreasing relevance). However, the temporal threshold is adjusted to 48 hours, as opposed to 1000 days in the e-commerce domain, based on the empirical findings of Li et al. [7].

For each article i, the temporal span τ_{ui} for the interaction with user u is defined as below,

$$\tau_{ui} = t_{ui} - pub_i \tag{2}$$

which represents the age of the article at the time the interaction occurred. t_{ui} is the time when user u interacted with item $\boldsymbol{i}$ (news click time), and pub_i indicates the publication time of item $\boldsymbol{i}$ (e.g., when the news was published). Articles are classified into temporal stages based on the threshold $T_{hot} = 48$ hours as below,

$$stage_i = \begin{cases} HOT & \text{if } \tau_{ui} \le T_{hot} \\ COLD & \text{if } \tau_{ui} > T_{hot} \end{cases} \tag{3}$$

The rationale for this classification is as follows. Articles in the HOT stage possess high intrinsic relevance and have likely been exposed to many users. Therefore, if a HOT article is not clicked, this constitutes a strong signal that the article is not relevant to the user (true negative). On the other hand, the articles in the COLD stage have decreasing relevance. The reason for the lack of clicks may be that the article is old or that the data is missing. This distinction is important in deciding the weights for the non-clicked articles.

3.3 Differential weighting strategy

The differential weighting method used in PA-TimeMF corrects for clicked items using position bias and for non-clicked items using temporal dynamics, without interference between the two corrections.

3.3.1 Position bias correction for clicked items

Position bias causes click for top-positioned items to be disproportionately higher than their actual relevance [15], leading the raw click signal to overestimate item relevance. For clicked items where r_{ui} = 1, the IPS corrects this by reweighting each interaction by $1/\theta_k$, up-weighting lower positions and down-weighting top positions as below,

$$w_{ui}^{click} = \frac{1}{\theta_k} \tag{4}$$

where, θ_k is the examination probability at position k estimated using the EM algorithm by Wang et al. [6].

3.3.2 Expectation-maximization algorithm for position bias estimation

E-step: Given the current estimate $\theta^{(t)}$, the expected relevance is calculated as below,

$$rel_{ui}^{(t)} = \frac{r_{ui}}{\theta_{p_{ui}}^{(t)}} \tag{5}$$

M-step: Examination probability for each position k is updated as below,

$$\theta_k^{(t+1)} = \frac{\sum_{(u,i):p_{ui}=k} r_{ui}}{\sum_{(u,i):p_{ui}=k} rel_{ui}^{(t)}} \tag{6}$$

The algorithm is initialized with uniform examination probabilities and iterated until convergence. The resulting value θ_k represents the position-dependent exposure probability and is used in IPS weighting.

3.3.3 Temporal weighting for non-clicked items

For non-clicked items ($r_{ui} = 0$), temporal weighting based on article stage is applied to distinguish true negatives from missing data. By adapting the dynamic popularity weighting from Li et al. [7], the weight is defined as below,

$$w_{ui}^{non\text{-}click} = w_0 \cdot \frac{pop_i^{sig(stage_i)\cdot\alpha}}{\sum_{j=1}^{N} pop_j^{sig(stage_j)\cdot\alpha}} \tag{7}$$

where, w_0 is the base weight for non-clicked items (hyperparameter), pop_i is the popularity of article i (measured as the total number of clicks), $\alpha \in [0,1]$ exponent that controls the degree of popularity influence, $sig(stage_i)$ is a function based on temporal stage:

$$sig(stage_i) = \begin{cases} +1 & \text{if } stage_i = HOT \\ -1 & \text{if } stage_i = COLD \end{cases} \tag{8}$$

α is used to determine the sensitivity of the weighting scheme to article popularity. Without α, all non-clicked articles would receive equal weight regardless of popularity, making it impossible to distinguish a truly irrelevant article from one that was simply not exposed. By introducing α as the exponent, when sig = +1 (HOT), the weight of popular non-clicked articles is amplified as strong true negatives, while when sig = −1 (COLD), the weight is suppressed as non-clicks likely reflect temporal irrelevance rather than genuine user disinterest.

3.3.4 Combined weighting scheme

The combined weighting scheme is formulated as below,

$$w_{ui} = \begin{cases} \dfrac{1}{\theta_{p_{ui}}} & \text{if } r_{ui} = 1 \text{ (click)} \\ w_0 \cdot \dfrac{pop_i^{sig(stage_i)\cdot\alpha}}{\sum_{j=1}^{N} pop_j^{sig(stage_j)\cdot\alpha}} & \text{if } r_{ui} = 0 \text{ (non-click)} \end{cases} \tag{9}$$

The main innovation of this approach is that it separates corrections, where IPS weighting of clicked items does not interfere with temporal correction of non-clicked items, which avoids compound errors that are seen with sequential approaches [8].

3.4 Unified objective function

The objective function for PA-TimeMF integrates position bias correction and temporal dynamics modeling in one unified formulation based on the weighted matrix factorization framework [14]:

$$\mathcal{L} = \sum_{u=1}^{M} \sum_{i=1}^{N} w_{ui} \cdot (r_{ui} - \mathrm{p}_u^T \mathrm{q}_i)^2 + \lambda\left(\sum_{u=1}^{M} \| \mathrm{p}_u \|^2 + \sum_{i=1}^{N} \| \mathrm{q}_i \|^2\right) \tag{10}$$

where, λ is the regularization parameter for preventing overfitting. This objective function has two primary components, which are described in detail below.

3.4.1 Weighted reconstruction error

$$\mathcal{L}_{data} = \sum_{u,i:r_{ui}=1} \frac{1}{\theta_{p_{ui}}} (1 - \mathrm{p}_u^T \mathrm{q}_i)^2 + \sum_{u,i:r_{ui}=0} w_{ui}^{non\text{-}click} (\mathrm{p}_u^T \mathrm{q}_i)^2 \tag{11}$$

The first component is in charge of correcting the overestimation of clicked items at top positions through IPS weighting. On the other hand, the second component is responsible for determining the weights of non-clicked items in accordance with their temporal stage and popularity level.

3.4.2 Regularization term

$$\mathcal{L}_{reg} = \lambda\left(\sum_{u=1}^{M} \| p_u \|^2 + \sum_{i=1}^{N} \| q_i \|^2\right) \tag{12}$$

This regularization component attempts to ensure that parameters are not excessively large in scale by penalizing model complexity, thereby improving generalization ability.

3.5 Optimization algorithm

The optimization process is carried out using Adaptive Gradient Descent (AdaGrad) optimization as shown in Algorithm 1. Unlike conventional gradient descent, AdaGrad dynamically adjusts the learning rate for each parameter based on the accumulation of gradients. This mechanism allows frequently updated parameters to have smaller update steps, while infrequently updated parameters retain relatively large learning rates, making it highly effective for handling sparse data in implicit feedback.

Algorithm 1. Position-aware temporal matrix factorization (PA-TimeMF) training

Input: Interaction matrix R, positions P, timestamps T, publication times $pub,$ latent dimension d, regularization parameter λ, base weight w_0, exponent α
Output: User factor matrix p, item factor matrix q

```
Initialize p_u, q_i ~ N(0, 0.01)
// Pre-computation
// Expectation-Maximization (EM) Algorithm
Repeat
 // E-step
 for each (u,i) do [Eq. (5)]
 // M-step
 for each position k do [Eq. (6)]
until convergence
//Classify temporal
for each article i do [Eqs. (2)–(3)]
// popularity count
pop_i, = ∑_u r_ui for all i
// Compute weights
for each (u,i) do
 if r_ui = 1 (click) do [Eq. (4)]
 if r_ui = 0 (non-click) do [Eq. (7)–(8)]
 // Optimize
 repeat (max 30 epochs)
   ℒ ← Weighted_MF_Loss (R, p, q, w, λ) [Eq. (10)]
   Update {p_u} using AdaGrad on ∂ℒ/∂p_u
   Update {q_i} using AdaGrad on ∂ℒ/∂q_i
 until convergence (ε = 10^-4)
 return {p_u}, {q_i }
```

Convergence Criteria:
The training process is terminated when the relative change in the objective function is less than the threshold $\varepsilon = 10^{-4}$ or after a maximum of 30 epochs, following standard practice in matrix factorization [28].

3.6 Ranking process and top-*K* selection

After the training process is completed, the preference score of user *u* for article *i* is predicted as:

$$\hat{r}_{ui} = p_u^T q_i \tag{13}$$

To generate top-K recommendations for user u, the following steps are performed:

1. Calculate prediction scores for all articles that have not been clicked by user u;
2. Sort articles by prediction score in descending order;
3. Return the K articles with the highest scores as recommendation results.

This framework allows for efficient online recommendation. Once the latent factors are learned, prediction is simply the computation of the inner product.

4. RESULT AND DISCUSSION

4.1 Experimental setup

4.1.1 Dataset

The experiments were conducted on a sample of data from the Ekstra Bladet News Recommendation Dataset, which comprises 1,000 unique users, 860 news articles, and 16,492 browsing sessions within 6 days (May 25 – June 1, 2023). EB-NeRD was selected as the most suitable dataset for PA-TimeMF's evaluation. It provides complete position metadata (article_ids_inview) for EM-based examination probability estimation and publication timestamps (published time) for HOT/COLD temporal classification, both essential to PA-TimeMF's methodology. While larger datasets such as MIND offer broader user coverage, they lack impression-level position metadata, limiting their suitability for position bias estimation. The dataset includes 132,876 total interactions, with 14,007 clicks (10.54% CTR). After filtering articles with fewer than 5 impressions, the final dataset comprises 792 articles, 999 users, and 104,787 interactions, with a CTR of 12.28% and sparsity of 87.72%. Although the observation window spans only 6 days, this is sufficient for PA-TimeMF's evaluation. The model's temporal classification operates on a 48-hour lifecycle threshold, meaning 6 days captures approximately three complete article lifecycle cycles. The CTR of 12.28% is also consistent with reported ranges in news recommendation literature (8–15%), confirming realistic user behavior.

4.1.2 Train/test split

In this study, the data is split using a temporal split with a ratio of 80:20, where 83,829 interactions are used to train the model, and 20,958 interactions are used to test the model. This is to ensure that the model is tested based on its ability to predict the future clicks.

4.1.3 Evaluation metrics

Following standard evaluation practices for implicit feedback-based recommendation systems [14], the metrics used include: (1) HR@10 (Hit Rate) as the primary metric in accordance with best practices in recommendation, (2) NDCG@10 for position-aware evaluation, (3) MAP@10 and MRR@10 as additional metrics for ranking quality.

4.1.4 Hyperparameter

Hyperparameter tuning was conducted using random search with 10 parameter combinations evaluated on the EB-NeRD dataset for 30 epochs per configuration. The optimized parameters include latent dimension ($K \in \{32, 64, 128\}$), regularization coefficient ($\lambda \in \{0.001, 0.01, 0.1\}$), temporal parameter $d_0 \in \{128, 256, 512\}$, and decay factor $\alpha \in \{0.2, 0.4, 0.6, 0.8\}$. The tuning results showed optimal configuration at $K = 64$, $\lambda = 0.01$, $d_0 = 256$, and $\alpha = 0.4$ with NDCG@10 performance of 0.0187.

4.2 Exploratory data analysis

4.2.1 Position bias

Position bias represents one of the primary sources of bias in ranking-based recommendation systems, as user behavior is heavily influenced by the presentation order of items. To identify the presence and extent of position bias in the dataset, an exploratory analysis was conducted on the distribution of CTR based on recommendation position. The exploration results can be seen Table 1.

Table 1. Position bias analysis based on CTR distribution and examination probability (θ_k) at each recommendation position.

Position	Impressions	Clicks	CTR (%)	θ_k	IPS Weight
1	16,492	1,893	11.48	0.1425	7.02
2	16,492	1,981	12.01	0.1437	6.96
3	16,492	1,908	11.57	0.1335	7.49
4	16,492	1,946	11.80	0.1278	7.83
5	16,492	1,959	11.88	0.1260	7.93
6	13,700	1,441	10.52	0.1012	9.88
7	11,492	1,029	8.95	0.0933	10.72
8	9,742	834	8.56	0.0771	12.96
9	8,363	557	6.66	0.0726	13.78
10	7,119	459	6.45	0.0713	14.03

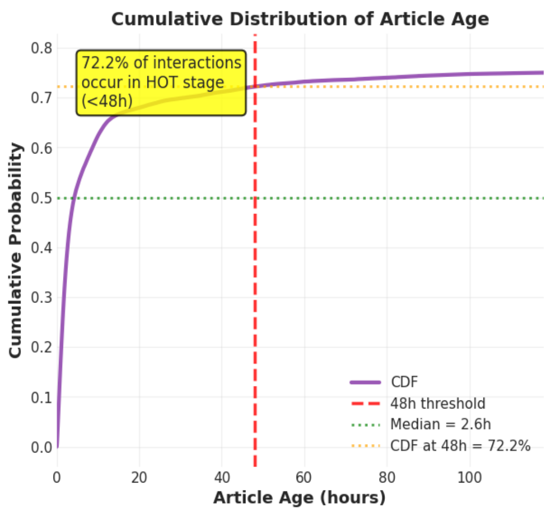
Note: CTR = lick-through rate, IPS = Inverse Propensity Score.

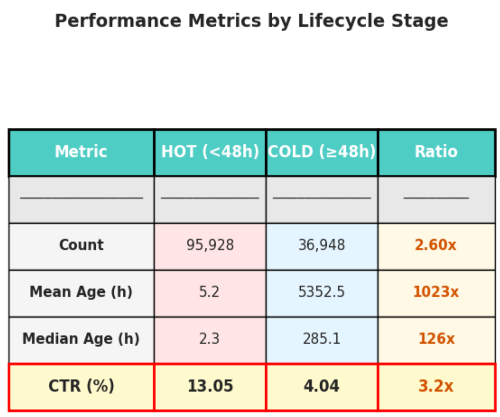
As shown in Table 1, the CTR at position 1 is 11.48%, gradually decreasing to 6.45% at position 10, with a position bias ratio of 1.78×, clearly indicating significant position bias. This is close to Collins et al. [5] observation that top positions receive significantly more clicks than lower positions, which confirms the need for position bias correction. The estimation of examination probability by applying the EM algorithm also indicates a similar bias, where $\theta_1 = 0{,}1425$ and $\theta_{10} = 0{,}0713$ with a ratio of 2.0×, meaning that items located at the first position are twice as likely to be examined as those located at the tenth position.

4.2.2 Temporal dynamics

Temporal dynamics in news recommendation domain are reflected in the distribution of article age throughout its lifecycle. This pattern reflects the rate at which article relevance decreases after publication as well as the concentration of user interaction intensity within specific time periods. Figure 2 illustrates the distribution of article age along with the lifecycle phase divisions applied to the analysis of the temporal characteristics of news content.

Figure 2 illustrates the rapid decay pattern, where the median equals 2.6 hours, 75th percentile equals 8.3 hours, and 95th percentile equals 118.0 hours or 4.9 days. Most importantly, 72.2% of the interactions involve articles that are less than 48 hours. This finding validates the selection of the 48-hour threshold adapted from Li et al. [7], who used a 1,000-day threshold in the e-commerce domain.





Performance Metrics by Lifecycle Stage

Metric	HOT (<48h)	COLD (≥48h)	Ratio
Count	95,928	36,948	2.60x
Mean Age (h)	5.2	5352.5	1023x
Median Age (h)	2.3	285.1	126x
CTR (%)	13.05	4.04	3.2x

Figure 2. Analysis of temporal characteristics of articles and justification for the 48-hour lifecycle threshold

The difference of CTR between HOT and COLD articles is highly significant. HOT articles have a CTR of 13.05%, while COLD articles have only a CTR of 4.04%, with an absolute difference of 9.01 percentage points and a relative ratio of 3.2×. This result is consistent with An et al. [4], which stated that 70-85% of clicks happen in the first 48 hours in the MIND dataset. This consistency provides strong empirical evidence for applying threshold-based temporal dynamics modeling in the news recommendation system field.

4.3 Position bias correction strategies

To determine the most effective position bias correction method used in PA-TimeMF, a comparison of five different approaches was conducted, including those that do not use debiasing, propensity-based approaches, and those that use relevance and position bias separation architectures. These approaches include EM + IPS (Ours), Baseline MF, Propensity Independent, Doubly Robust, and Two-Tower Revisited. Since PA-TimeMF operates within a matrix factorization framework, all baselines were adapted to the same setting for a controlled comparison of position bias correction principles. From Two-Tower, only the position bias separation principle was extracted and implemented within the MF framework, replacing the neural dual-encoder with latent factor matrices. From Doubly Robust, only the doubly robust loss estimator was extracted, with EM-estimated propensity scores replacing the original propensity network. This ensures that all methods are evaluated under identical conditions, using the same dataset, MF backbone, and evaluation metrics, isolating the contribution of each debiasing principle independently of architectural differences. The evaluation results of these approaches are shown in Figure 3.

As indicated in the experimental results, it is clear that the best performance is achieved by the proposed EM+IPS method on all the evaluation metrics in comparison to the other methods. For instance, as shown in Figure 3, the performance of the proposed method in terms of HR@10, NDCG@10, MAP@10, and MRR@10 is 0.100, 0.023, 0.012, and 0.037, respectively.

In comparison to the Baseline MF method, the proposed EM+IPS method shows substantial relative improvements across all evaluation metrics, as shown in Figure 3. Specifically, the improvements are +32.7% in HR@10, +42.8% in NDCG@10, +57.0% in MAP@10, and +38.9% in MRR@10.

The analysis of the performance of the other methods shows surprising results, as all the state-of-the-art methods show performance degradation in comparison to the Baseline MF method. For instance, the Propensity-Independent method shows a degradation of −24.9% in HR@10 and −38.4% in NDCG@10, while the Doubly Robust method shows a drastic degradation of −38.4% in HR@10 and −24.9% in MAP@10.

This trend of performance also shows that the mentioned approaches are not applicable to the news recommendation domain with its prominent temporal characteristics. The Neural-based approaches, such as Two-Tower and Doubly Robust, show the tendency to over-correct the position bias, thus resulting in underfitting models in comparison to actual user interaction patterns. This performance degradation suggests that their debiasing mechanisms, when stripped of their original architectural context, tend to over-correct position bias in the MF setting. Propensity-Independent, on the contrary, does not use propensity scores, thus failing to incorporate important positional information necessary for proper debiasing.

The results of the comparison validate the choice of the position bias correction component in the PA-TimeMF model, namely, EM+IPS. The inability of the neural-based approaches to outperform the baseline also emphasizes the importance of understanding the limitations of approaches that were proven to be effective in the e-commerce and general recommendation domains, and the specific characteristics of the news recommendation domain.

Figure 3. Comparison of position bias correction methods in position-aware temporal matrix factorization (PA-TimeMF) and relative improvement over baseline

4.4 Temporal dynamics modeling strategies

To assess the effectiveness of the temporal part in PA-TimeMF, five different temporal modeling strategies are considered: TimeMF (Ours), LANCER, Baseline MF, HyperNES, and PP-Rec. For each of the baseline models, only the temporal part is used, and the other parts of the original models are not considered to conduct a fair comparison. The experimental results are shown in Figure 4.

From the experimental results, TimeMF (Ours) performs better than the other models in all aspects of the evaluation metrics. As shown in Figure 4, TimeMF achieves an HR@10 of 0.091, NDCG@10 of 0.021, MAP@10 of 0.011, and MRR@10 of 0.033. These results show a significant improvement over the other models. Compared to Baseline MF, TimeMF achieves relative improvements of +31.2% in HR@10, +34.6% in NDCG@10, +39.9% in MAP@10, and +22.1% in MRR@10, respectively, as shown in Figure 4. These improvements in all aspects of the evaluation metrics indicate that the modeling of temporal dynamics is an important aspect of improving the accuracy of news recommendations.

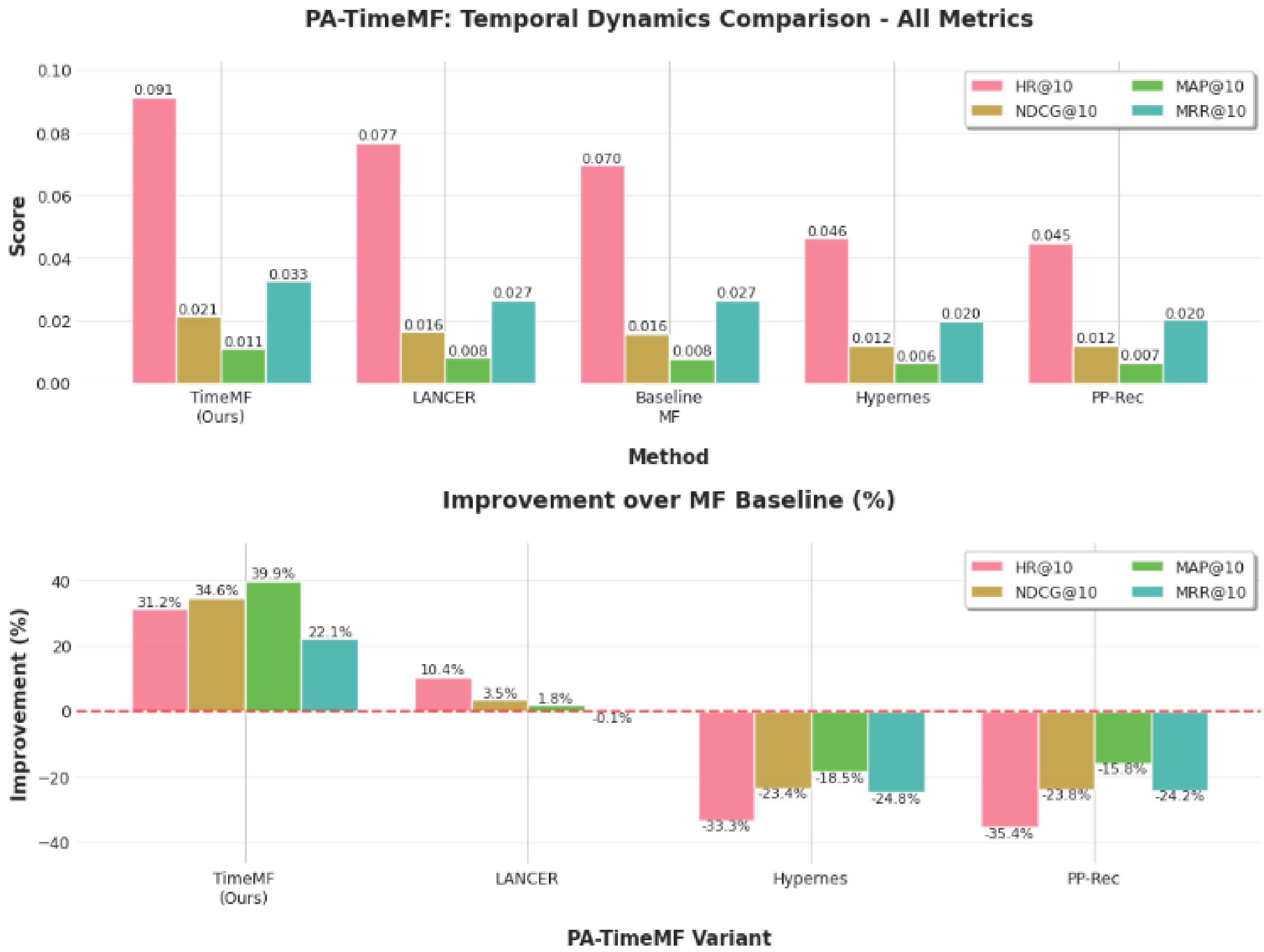


Figure 4. Comparison of temporal dynamics modeling strategies in position-aware temporal matrix factorization (PA-TimeMF) and relative improvement over baseline

The second-best performance is shown in LANCER, which has moderate improvements over Baseline MF: +10.4% in HR@10, +3.5% in NDCG@10, and +1.8% in MAP@10. Although LANCER uses more advanced adaptive decay rate learning, the improvements are much lower compared to TimeMF. This shows that the additional complexity of LANCER does not translate to improvements commensurate with the additional computational cost.

On the other hand, the performance of HyperNES and PP-Rec deteriorated significantly. HyperNES showed droped of −33.3% in HR@10, −23.4% in NDCG@10, and −18.5% in MAP@10, while PP-Rec drops −35.4% in HR@10, −23.8% in NDCG@10, and −24.2% in MRR@10.

These results show that lightweight lifecycle-aware temporal modeling outperforms both neural and popularity-based approaches. This is consistent with the news domain's characteristics, namely predictable temporal decay patterns, clear HOT/COLD lifecycle boundaries, and more informative collaborative signals for personalization.

4.5 Ablation study: Component contributions

In this section, we carry out an ablation study to verify the effectiveness of each component of the proposed PA-TimeMF method. To perform the ablation study, we compare four versions of the model: (1) Baseline MF without any additional components, (2) MF + Position, where we use position bias correction as the additional component, (3) MF + Temporal, where we use temporal dynamics as the additional component, and (4) PA-TimeMF (Full), where we use the combined effect of both components with the proposed unified learning framework. The complete ablation study results are shown Figure 5 and Table 2.

The experimental results verify that PA-TimeMF (Full) performs the best in all evaluation metrics. All metrics are reported as mean ± standard deviation computed across users in the test set. The small standard deviations across all methods confirm the stability and consistency of the reported improvements. Statistical significance of the performance differences is further validated through paired *t*-test and Cohen's d effect size analysis, as presented in Table 3. As shown in Figure 5 and Table 2, the HR@10, NDCG@10, MAP@10, and MRR@10 of PA-TimeMF are 0.1188, 0.0258, 0.0132, and 0.0389, respectively. Compared with Baseline MF, the relative improvements of PA-TimeMF are +57.7% in HR@10, +58.7% in NDCG@10, +68.6% in MAP@10, and +46.9% in MRR@10.

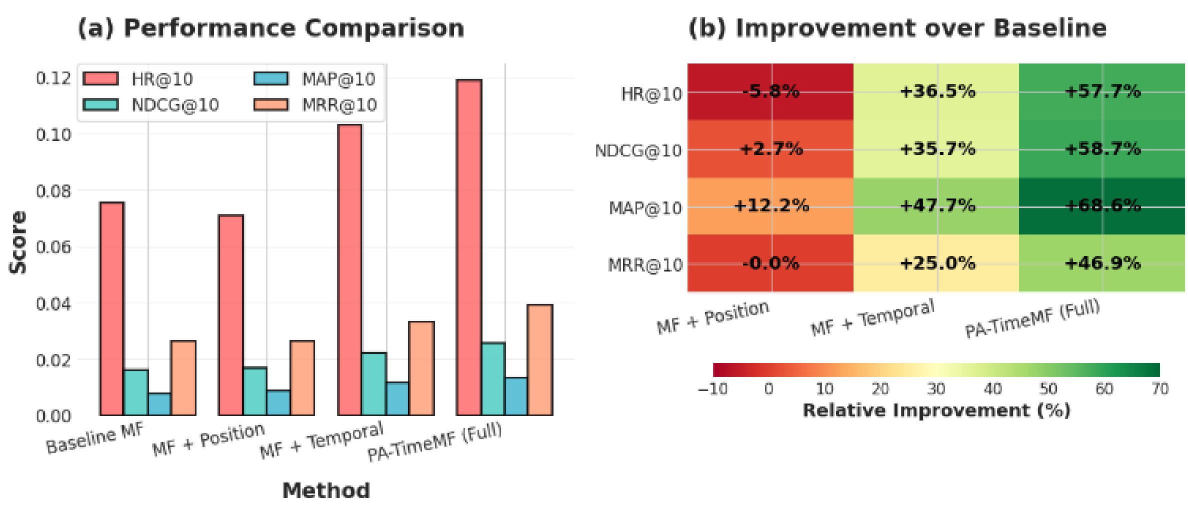


Figure 5. position-aware temporal matrix factorization (PA-TimeMF) ablation study: performance comparison and model component contributions

Table 2. PA-TimeMF ablation study results across various evaluation metrics (mean ± std across users; relative improvements over baseline mf in parentheses)

Method	HR@K	NDCG@K	MAP@K	MRR@K
Baseline MF	0.0754 ± 0.0018	0.0163 ± 0.0005	0.0078 ± 0.0003	0.0265 ± 0.0009
MF + Position	0.0710 ± 0.0016	0.0167 ± 0.0005	0.0088 ± 0.0003	0.0265 ± 0.0008
MF + Temporal	0.1029 ± 0.0021	0.0221 ± 0.0006	0.0115 ± 0.0004	0.0331 ± 0.0011
PA-TimeMF	0.1188 ± 0.0019	0.0258 ± 0.0006	0.0132 ± 0.0003	0.0389 ± 0.0010

Note: PA-TimeMF = position-aware temporal matrix factorization.

The component contribution results show that the temporal dynamics factor is the most important factor in improving performance. The MF + Temporal model obtains large improvements of +36.5% in HR@10 and +35.7% in NDCG@10, and consistent improvements in all metrics. This verifies the importance of temporal dynamics in the highly time-sensitive news domain. On the contrary, the MF + Position model shows a degradation in HR@10 of -5.8% and limited improvements in MAP@10 of +12.2%. This shows that position bias correction may lead to over-correction and loss of relevance in articles that are very popular in the dataset when the temporal effect is not considered.

The integration of both components in PA-TimeMF also results in synergistic effects, where the combined effects of both components in the model exceed the individual effects of each component. This is confirmed in Figure 5, where the effect of using the PA-TimeMF model is much higher than the combined effects of the MF + Temporal and MF + Position models, thus confirming the interaction between position bias and temporal dynamics, which results in confounding effects.

Statistical validation was also conducted to assess the significance of performance differences among model variants. The results of the statistical validation test are presented in Figure 6 and Table 3.

Statistical validation by paired *t*-test and Cohen's *d* effect size measure further reinforces these empirical results. Cohen's *d* is used alongside the paired *t*-test to measure practical significance beyond statistical significance alone, following Cohen's conventional criteria: small ($d < 0.2$), medium ($0.2 \leq d < 0.5$), and large ($d \geq 0.5$). As illustrated in Figure 4 and Table 3, the difference between PA-TimeMF and Baseline MF has a Cohen's $d = 0.120$ with $p < 0.01$, which is highly statistically significant with a small effect size. The difference between PA-TimeMF and MF + Position is also statistically significant (Cohen's $d = 0.110$, $p < 0.05$), thus confirming that the positive contribution of position bias correction is achieved when combined correctly with temporal modeling. On the other hand, the difference between PA-TimeMF and MF + Temporal is not statistically significant (Cohen's $d = 0.043$, $p = 0.1046$), which suggests that adding the position bias component to a model that already incorporates temporal dynamics results in marginal but consistent improvements. Although all effect sizes are small by conventional standards, the statistical significance of key comparisons ($p < 0.01$ and $p < 0.05$) confirms that the observed improvements are consistent and reliable, rather than due to random variation. In news recommendation, consistent improvements in top-k ranking quality translate to meaningful gains in user experience even when effect sizes are modest.

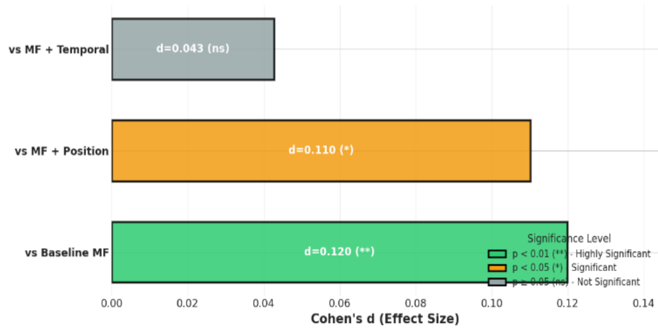


Figure 6. Visualization of statistical significance test results and effect sizes among model variants

Table 3. Statistical significance test results among model variants

Method	*t*-Statistic	*p*-Value	Cohen's *d*	Sig.
Baseline MF	2.67	0.0078	0.120	**
MF + Position	2.47	0.0136	0.110	*
MF + Temporal	1.63	0.1046	0.043	ns

Note: MF: matrix factorization.

From the ablation study findings, three key findings can be concluded. Firstly, the temporal dynamics play the major role in contributing the most towards the performance improvements. Secondly, position bias correction is complementary in nature. Applied in isolation it degrades performance, but combined with temporal dynamics it yields positive gains. Thirdly, their joint application produces a synergistic effect, validating PA-TimeMF's architectural design. Overall, the findings from this ablation study confirm that the application of bias correction strategies in isolation distorts the relevance signals of the implicit feedback.

4.6 Comparison with state-of-the-art methods (matrix factorization-based)

To situate PA-TimeMF in the context of existing news recommendation approaches, we compared PA-TimeMF with three categories of baselines developed in the matrix factorization paradigm: popularity-based (GLORY-inspired), content-based (NRMS-inspired), and temporal user modeling (LSTUR-inspired). All these baselines were modified from neural SOTA models to the collaborative filtering setting to make a fair comparison with PA-TimeMF. The comparison results are shown in Table 4.

Table 4. Performance comparison of PA-TimeMF with SOTA baselines adapted to the matrix factorization framework

Method	HR@10	NDCG@10	MAP@10	MRR@10
PA-TimeMF (Ours)	0.1000	0.0232	0.0123	0.0368
Popularity (GLORY)	0.0812	0.0189	0.0095	0.0307
Content-based (NRMS)	0.0754	0.0161	0.0079	0.0267
Temporal User (LSTUR)	0.0725	0.0150	0.0072	0.0260

Note: PA-TimeMF = position-aware temporal matrix factorization.

The experimental results show that PA-TimeMF performs better than all other baseline approaches on all evaluation metrics. As shown in Table 4, PA-TimeMF achieves the highest HR@10 of 0.1000, NDCG@10 of 0.0232, MAP@10 of 0.0123, and MRR@10 of 0.0368, outperforming all other approaches in terms of accuracy and ranking quality.

In comparison with GLORY (popularity-based), PA-TimeMF achieves relative improvements of +19.9% to +29.5%. This suggests that although popularity information is valuable, models that neglect position bias correction and temporal modeling may generate less personalized recommendations and are suboptimal in ranking quality.

When compared with NRMS (content-based), the improvement is even higher, ranging from +32.6% to +55.7%. This suggests that content similarity alone is not sufficient in the news recommendation domain. Content-based features capture semantic relevance but neglect consumption timing and position effects, both of which significantly impact click behavior.

The best performance is achieved when compared with LSTUR (temporal user modeling), with a performance difference ranging from +37.9% to +70.8%. This confirms that item-side temporal decay (content freshness) is more significant than user-side temporal decay (preference changes) in the news domain. These results collectively confirm the effectiveness of PA-TimeMF's integrated approach.

4.7 Discussion

The experimental results verify that position bias and temporal dynamics are not independent in news recommendation systems. From the ablation study, it can be seen that the performance drops by −5.8% HR@10 when only correcting the position bias individually, while the performance increases by +57.7% HR@10 when integrating with temporal dynamics. This phenomenon reveals that there is a confounding effect when new articles are ranked higher and gain the maximum exposure.

The superiority of temporal dynamics (+36.5% HR@10) compared with correcting the position bias verifies the special characteristics of the news domain, where the temporal dynamics are the most important factor. This is also verified by An et al. [4], where they found that 70–85% clicks happen within the first 48 hours; however, it is also important to note that the temporal dynamics and position bias correction need to be integrated to obtain the optimal performance.

Comparison with SOTA-inspired approaches gives valuable information about task-agnostic method design. SOTA-inspired approaches like Doubly Robust and Two-Tower observe a decline in performance (−13.5% to −38.4%) when modified to the collaborative filtering paradigm, suggesting that architectural complexity is not necessarily beneficial without bias and temporal modeling in the news setting. In this scenario, PA-TimeMF's lightweight design with bias modeling is more stable and effective.

The biggest difference in performance against LSTUR (+70.8% MAP@10) highlights the paradigmatic shift from user-centric to item-centric temporal modeling. In the e-commerce setting, the evolution of user preferences is a relatively slow process, and thus user-side temporal modeling is more effective. In contrast, news articles decay rapidly in relevance (hours/days), and thus item-side temporal modeling is more important. PA-TimeMF's lifecycle-aware article decay modeling is therefore more appropriate for the news domain.

However, there are certain limitations. Firstly, this research is evaluated on a single dataset (EB-NeRD) with 792 articles and 999 users. Although EB-NeRD is the most suitable available dataset for PA-TimeMF's methodology, its limited scale and single-platform nature restrict generalizability to cross-platform or multilingual news settings. Additionally, the 6-day observation window, while sufficient for the 48-hour lifecycle threshold, may not capture long-term behavioral patterns such as seasonal trends or evolving user interests. Cross-dataset validation and longer observation periods remain important directions for future work. Secondly, the lifecycle threshold is fixed at 48 hours and the temporal classification is limited to a binary HOT/COLD scheme. Both the threshold value and granularity may vary for different news types. Thirdly, the cold start problem for both users and articles is not explicitly considered.

Future research directions include: (1) cross-dataset validation on larger and more diverse corpora; (2) learning adaptive thresholds from article metadata (e.g., category, breaking news flag) and modeling multi-granular temporal decay functions that vary per user or content category, replacing the fixed binary classification with dynamically learned boundaries; and (3) exploring hybrid and causal approaches for cold start and confounding bias measurement.

5. CONCLUSION

This study introduces PA-TimeMF, a news recommendation system that tackles position bias and temporal dynamics at the same time through a unified matrix factorization-based objective function. It incorporates Inverse Propensity Scoring using the EM algorithm for position bias mitigation and lifecycle-aware temporal weighting to distinguish between HOT items (<48 hours) and COLD items (≥48 hours).

The evaluation of the proposed PA-TimeMF on the EB-NeRD dataset shows three key findings. First, PA-TimeMF outperforms other methods in a controlled matrix factorization setup, obtaining improvements of up to 57.7% in HR@10, 58.7% in NDCG@10, and 68.6% in MAP@10 compared to the baseline method. Second, the ablation study shows that the synergistic effect of position bias mitigation and temporal dynamics is significant, where temporal dynamics play the dominant role and position bias mitigation makes complementary contributions. Third, comparing PA-TimeMF with SOTA-inspired approaches in the MF framework shows the effectiveness of the proposed approach, especially against temporal user modeling, proving that item-side temporal decay is more important than user-side temporal evolution in news recommendation.

Overall, this research demonstrates that position bias and temporal dynamics must be modeled simultaneously for optimal performance. A lightweight approach with explicit bias handling can achieve competitive results with lower computational overhead. Future work on adaptive temporal modeling and content integration is expected to further enhance recommendation robustness and personalization.

REFERENCES

[1] Newman, N., Fletcher, R., Robertson, C.T., Ross Arguedas, A., Nielsen, R.K. (2024). Reuters Institute digital news report 2024. Reuters Institute for the Study of Journalism. https://doi.org/10.60625/RISJ-VY6N-4V57

[2] Sonni, A.F. (2025). Digital transformation in journalism: Mini review on the impact of AI on journalistic practices. Frontiers in Communication, 10. https://doi.org/10.3389/fcomm.2025.1535156

[3] Wu-Ouyang, B., Lu, S., Chen, H.T. (2025). News avoidance or curation? Explicating the psychological process in news consumption on Facebook. New Media & Society. https://doi.org/10.1177/14614448251351280

[4] An, M., Wu, F., Wu, C., Zhang, K., Liu, Z., Xie, X. (2019). Neural news recommendation with long- and short-term user representations. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, pp. 336-345. https://doi.org/10.18653/v1/p19-1033

[5] Collins, A., Tkaczyk, D., Aizawa, A., Beel, J. (2018). Position bias in recommender systems for digital libraries. In Lecture Notes in Computer Science, Springer International Publishing, pp. 335-344. https://doi.org/10.1007/978-3-319-78105-1_37

[6] Wang, X., Bendersky, M., Metzler, D., Najork, M. (2016). Learning to rank with selection bias in personal search. In Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, Pisa, Italy, pp. 115-124. https://doi.org/10.1145/2911451.2911537

[7] Li, M., Wu, H., Zhang, H. (2019). Matrix factorization for personalized recommendation with implicit feedback and temporal information in social ecommerce networks. IEEE Access, 7: 141268-141276. https://doi.org/10.1109/access.2019.2943959

[8] Chen, J., Dong, H., Wang, X., Feng, F., Wang, M., He, X. (2023). Bias and debias in recommender system: A survey and future directions. ACM Transactions on Information Systems, 41(3): 1-39. https://doi.org/10.1145/3564284

[9] Zheng, Y., Gao, C., Li, X., He, X., Li, Y., Jin, D. (2021). Disentangling user interest and conformity for recommendation with causal embedding. In Proceedings of the Web Conference 2021, Ljubljana, Slovenia, pp. 2980-2991. https://doi.org/10.1145/3442381.3449788

[10] Yang, M., Dai, Q., Dong, Z., Chen, X., He, X., Wang, J. (2021). Top-N recommendation with counterfactual user preference simulation. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management, Virtual Event, Queensland Australia, pp. 2342-2351. https://doi.org/10.1145/3459637.3482305

[11] Chaney, A.J., Stewart, B.M., Engelhardt, B.E. (2018). How algorithmic confounding in recommendation systems increases homogeneity and decreases utility. In Proceedings of the 12th ACM Conference on Recommender Systems, Vancouver, British Columbia, Canada, pp. 224-232. https://doi.org/10.1145/3240323.3240370

[12] Mansoury, M., Abdollahpouri, H., Pechenizkiy, M., Mobasher, B., Burke, R. (2020). Feedback loop and bias amplification in recommender systems. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, Virtual Event, Ireland, pp. 2145-2148. https://doi.org/10.1145/3340531.3412152

[13] Hu, Y., Koren, Y., Volinsky, C. (2008). Collaborative filtering for implicit feedback datasets. In 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, pp. 263-272. https://doi.org/10.1109/icdm.2008.22

[14] Chen, J., Zhang, H., He, X., Nie, L., Liu, W., Chua, T.S. (2017). Attentive collaborative filtering. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan, pp. 335-344. https://doi.org/10.1145/3077136.3080797

[15] Wu, C., Wu, F., Ge, S., Qi, T., Huang, Y., Xie, X. (2019). Neural news recommendation with multi-head self-attention. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, pp. 6388-6393. https://doi.org/10.18653/v1/d19-1671

[16] Joachims, T., Swaminathan, A., Schnabel, T. (2017). Unbiased learning-to-rank with biased feedback. In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, Cambridge, United Kingdom, pp. 781-789. https://doi.org/10.1145/3018661.3018699

[17] Agarwal, A., Zaitsev, I., Wang, X., Li, C., Najork, M., Joachims, T. (2019). Estimating position bias without intrusive interventions. In Proceedings of the Twelfth

ACM International Conference on Web Search and Data Mining, Melbourne, VIC, Australia, pp. 474-482. https://doi.org/10.1145/3289600.3291017
[18] Oosterhuis, H., Jagerman, R., de Rijke, M. (2020). Unbiased learning to rank: Counterfactual and online approaches. In Companion Proceedings of the Web Conference 2020, Taipei, Taiwan, pp. 299-300. https://doi.org/10.1145/3366424.3383107
[19] Bito, E., Ren, Y., He, E. (2025). Evaluating position bias in large language model recommendations. arXiv preprint arXiv:2508.02020. https://doi.org/10.48550/ARXIV.2508.02020
[20] Zhang, Q., Lu, J., Jin, Y. (2020). Artificial intelligence in recommender systems. Complex & Intelligent Systems, 7(1): 439-457. https://doi.org/10.1007/s40747-020-00212-w
[21] Li, J., Wang, Y., McAuley, J. (2020). Time interval aware self-attention for sequential recommendation. In Proceedings of the 13th International Conference on Web Search and Data Mining, Houston, TX, United States, pp. 322-330. https://doi.org/10.1145/3336191.3371786
[22] Ryu, S., Ko, Y., Kim, S.W. (2025). Is this news still interesting to you? Lifetime-aware interest matching for news recommendation. In Proceedings of the 34th ACM International Conference on Information and Knowledge Management, Seoul, Republic of Korea, pp. 2515-2524. https://doi.org/10.1145/3746252.3761047
[23] Iwai, T., Tomita, A., Arai, T., Ogawa, H., Saito, T. (2024). Harnessing temporal dynamics and content: An ensemble of gradient boosting machines for news recommendation. In Proceedings of the Recommender Systems Challenge 2024, Bari, Italy, pp. 37-41. https://doi.org/10.1145/3687151.3687159
[24] Qi, T., Wu, F., Wu, C., Huang, Y., Xie, X. (2020). Privacy-preserving news recommendation model learning. In Findings of the Association for Computational Linguistics: EMNLP 2020, pp. 1423-1432. https://doi.org/10.18653/v1/2020.findings-emnlp.128
[25] Guo, H., Tang, R., Ye, Y., Li, Z., He, X. (2017). DeepFM: A factorization-machine based neural network for CTR prediction. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, Melbourne, VIC, Australia, pp. 1725-1731. https://doi.org/10.24963/ijcai.2017/239
[26] He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., Wang, M. (2020). LightGCN. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, China, pp. 639-648. https://doi.org/10.1145/3397271.3401063
[27] Ekstrand, M.D., Burke, R., Diaz, F. (2019). Fairness and discrimination in retrieval and recommendation. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, Paris, France, pp. 1403-1404. https://doi.org/10.1145/3331184.3331380
[28] Rendle, S., Zhang, L., Koren, Y. (2019). On the difficulty of evaluating baselines: A study on recommender systems. arXiv preprint arXiv:1905.01395. https://doi.org/10.48550/arXiv.1905.01395
[29] Liu, D., Cheng, P., Zhu, H., et al. (2021). Mitigating confounding bias in recommendation via information bottleneck. In Fifteenth ACM Conference on Recommender Systems, Amsterdam, Netherlands, pp. 351-360. https://doi.org/10.1145/3460231.3474263
[30] Chen, Y., Ye, W., Lin, C., Chen, Y. (2024). Unbiased news recommendation model combining time and content. Expert Systems with Applications, 256: 124864. https://doi.org/10.1016/j.eswa.2024.124864
[31] Deldjoo, Y., Jannach, D., Bellogin, A., Difonzo, A., Zanzonelli, D. (2023). Fairness in recommender systems: Research landscape and future directions. User Modeling and User-Adapted Interaction, 34(1): 59-108. https://doi.org/10.1007/s11257-023-09364-z