Large Scale Hierarchical User Interest Modeling for Click-through Rate Prediction

Taofeng Xue
Meituan Search and Content Intelligence
China
xuetaofeng@meituan.com

Zhimin Lin
Meituan Search and Content Intelligence
China
linzhimin@meituan.com

Zijian Zhang
Meituan Search and Content Intelligence
China
zhangzijian14@meituan.com

Linsen Guo
Meituan Search and Content Intelligence
China
guolinsen@meituan.com

Haoru Chen
Meituan Search and Content Intelligence
China
chenhaoru02@meituan.com

Mengjiao Bao
Meituan Search and Content Intelligence
China
baomengjiao@meituan.com

Peng Yan
Meituan Search and Content Intelligence
China
yanpeng04@meituan.com

ABSTRACT

With the explosive growth of online information, recommender systems have emerged as indispensable tools for navigating the complexities of content generation, discovery and consumption. The RecSys 2024 Challenge, organized by Ekstra Bladet, provides a comprehensive dataset and a robust news recommendation evaluation framework for tackling multifaceted challenges including modeling user preferences based on implicit behavior, accounting for the influence of the news agenda on user interests, and managing the rapid decay of news items. In this paper, we propose a novel **H**ierarchical **U**ser **I**nterest **M**odeling (HUIM) approach to this challenge, leveraging both long-term invariant interests and short-term rapidly changing interests. Specifically, we utilize the multimodal representations along with the side information of items to distill the long-standing interests from user historical behaviors. By analyzing real-time context and user behavioral path patterns, we identify their fine-grained instant interests. Finally, an interest fusion network is proposed to adaptively fuse the long-term and short-term interests by contrasting the query-aware fine-grained behaviors with query-level cross entropy loss. Our team, BlackPearl, achieved a score of 0.8815 and ranked 2nd place on the final leaderboard.

CCS CONCEPTS

• **Information systems** → **Recommender systems**; **Content ranking**; **Data mining**.

KEYWORDS

Recsys Challenge, User Interest Modeling, Click-through Rate Prediction, Recommender Systems

RecSys Challenge '24, October 14–18, 2024, Bari, Italy

ACM ISBN 979-8-4007-1127-5/24/10
https://doi.org/10.1145/3687151.3687163

ACM Reference Format:
Taofeng Xue, Zhimin Lin, Zijian Zhang, Linsen Guo, Haoru Chen, Mengjiao Bao, and Peng Yan. 2024. Large Scale Hierarchical User Interest Modeling for Click-through Rate Prediction. In *ACM RecSys Challenge 2024 (RecSys Challenge '24), October 14–18, 2024, Bari, Italy*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3687151.3687163

1 INTRODUCTION

Recommender systems have seen a significant evolution over the past decades, transitioning from simple content-based filtering to personalized deep neural networks that adapt to users' interests in real time. Nevertheless, accurately predicting user click-through rates while embracing the normative complexities remains a key challenge for recommender systems.

The RecSys Challenge 2024 [1], organized by Ekstra Bladet, focuses on online news recommendation and provides a comprehensive dataset including over 2.7 million users and more than 600 million impression log. This challenge aims to address both the technical and normative challenges inherent in the design of effective and responsible recommender systems for news publishing. The goal is to rank the candidate articles based on the user's personal preferences, developing models that encapsulate both the users and the articles through their content and the users' interests.

In the news platform, users' reading behaviors are primarily influenced by two aspects. Firstly, users have their personal taste preferences, making the understanding and capture of their long-standing interests crucial. Furthermore, given the immediacy of news, fresh articles emerge daily, and breaking events can spark particular interest among users. How to consider both the long-term unchanging interests and the sudden interests of users is quite challenging. The DIN [7] and DIEN [6] leverage attention mechanisms to align user interests with item recommendations. TransAct [5] further enhances these capabilities by integrating multi-dimensional user engagement data. Additionally, the DropoutNet [3] addresses the cold start problem by employing dropout layers to manage uncertainty. MaskNet [4] and PEPNet [1] are capable of dynamically capturing the importance of embeddings. Despite their ability to capture some aspects of users' interests, these methods remain

[1] http://www.recsyschallenge.com/2024/

ineffective in tackling issues such as the drifting of user interests over time.

In this paper we present our approach to this challenge. We propose a hierarchical user interest model to effectively capture user diverse interests. The main contributions of our work are summarized as follow:

- We employ the multimodal representations along with the side information of items to distill the invariant interests from user historical behaviors and identify their fine-grained instant interests by delicated designed list-wise features.
- To enhance the recognition of user fine-grained interests, a interest fusion network is proposed to adaptively fuse the long-term and short-term interests by contrasting the query-aware fine-grained behaviors with query-level cross entropy loss.
- We conduct experiments on the dataset to demonstrate the effectiveness of our approach. Our team, BlackPearl, achieved a score of 0.873 for the single HUIM model, 0.8815 for the ensemble model and ranked 2nd place on the final leaderboard.

2 FORMULATION

In this section, we give a formal description of the dataset provided by Ekstra Bladet and formulate the problem of news recommendation. The main datasets include history data and impression log as well as the article metadata.

Historical Dataset $\mathcal{H} = \{\mathcal{H}_u\}$: For a user $u \in \mathcal{U}$, the history is characterized by a sequence of previously clicked articles, denoted as $\mathcal{H}_u = \{i_1, i_2, ..., i_m\}$ along with the corresponding timestamps of these interactions, represented by $\mathcal{T}_u = \{t_1, t_2, ..., t_m\}$. Additionally, the user's engagement with each article is captured by the read time sequence $R_u = \{r_1, r_2, ..., r_m\}$ and the scroll percentage sequence $S_u = \{s_1, s_2, ..., s_m\}$. Here, m represents the total number of historical interactions recorded for user u.

Impression Log $\mathcal{L} = \{(\mathcal{X}, y)\}$: Given the quadruplet $(u, q, p, i) \in \mathcal{X}$, which represents the exposure of article i to user u within the context of an impression identified by q and a session identified by p, a binary label $y \in \{0, 1\}$ denotes whether the article is clicked. It is important to note that a session is a coarser-grained request identifier than an impression. Within a single session, there may be multiple impressions, which represent the connections and transitions in the user real-time behavioral path.

Article Metadata $\mathcal{I}$: In addition to the regular auxiliary features such as category, title and topic, the organizers have also provided multimodal pre-trained embeddings for each article, i.e., $\mathcal{E} = \{e_1, e_2, ..., e_N\}$, where N denotes the total number of articles.

Problem (Click-through Rate Prediction for News Recommendation): Our task is to leverage the historical dataset $\mathcal{H}$ and the impression log $\mathcal{L}$ from both the training and validation datasets, as well as item metadata $\mathcal{I}$ to estimate the click-through rate for each item presented in the test dataset's impression log. The primary metric for the challenge is Area under the ROC curve (AUC) which is equivalent to the GAUC [7] defined at the impression level.

The temporal span relationships among the training dataset, validation dataset, and test dataset, as well as their corresponding historical behavioral datasets, are illustrated in the Figure 1.

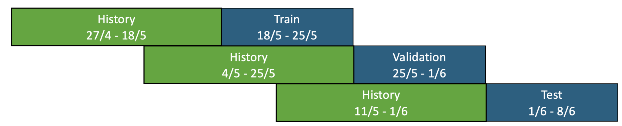

Figure 1: Temporal span relationships among the training dataset, validation dataset, and test dataset, as well as their corresponding historical behavioral datasets.

3 APPROACH: HUIM

In this section, we elaborate on the proposed **H**ierarchical **U**ser **I**nterest **M**odeling (HUIM) approach illustrated in Figure 2, detailing its structure and functionality to provide a comprehensive understanding of how it captures the multi-faceted dimensions of user interests within the intricate and dynamic real-world context.

Overall, we employ the multimodal representation and auxiliary information, enhanced by a target-aware mechanism to capture the user's historical interests, which reveals the relatively stable user preferences. In capturing instant interests, given the rapid changes in news that lead to cold start problems, we rely on refined feature engineering to deeply characterize the user's hierarchical behavior path patterns, detailing interests at the day granularity, session granularity, and impression granularity from coarse to fine. Ultimately, we adopt a interest fusion network to adaptively fuse the hierarchical interests by contrasting the query-aware fine-grained behaviors with query-level cross entropy loss. Finally, we achieved a score of 0.873 for the single HUIM model, 0.8815 for the blending model of HUIM and catboost and ranked 2nd place on the final leaderboard.

3.1 User Invariant Interest Modeling

we firstly discuss the representation of each article. Our statistics show that over 60% of the articles in the impression log do not appear in the historical behavior data. Hence, relying on article ids as a sole indicator is insufficient to effectively measure user long-standing interests. It is essential to utilize the auxiliary information and multimodal representation of the articles to characterize them and explore the underlying patterns in user historical invariant interests. Specifically, we incorporate key auxiliary information such as category, subcategory, sentiment label, and type to represent the target article.

$$\boldsymbol{e}_i = \text{Concat}(\boldsymbol{E}_i^{\text{category}}, \boldsymbol{E}_i^{\text{subcat}}, \boldsymbol{E}_i^{\text{sentiment}}, \boldsymbol{E}_i^{\text{type}}) \in \mathbb{R}^d \tag{1}$$

Specifically, for articles in historical behaviors, as the read time and scroll percentage provided for each article indicate the intensity of user engagement, we perform binning and discretization on them and concatenate their embeddings to $\boldsymbol{e}_i$ to form the article representation. This representation is then aligned with the dimensions of the target item using the matrix $\boldsymbol{W}^1$.

$$\boldsymbol{e}_j = \boldsymbol{W}^1\text{Concat}(\boldsymbol{E}_i^{\text{category}}, \boldsymbol{E}_i^{\text{subcat}}, \boldsymbol{E}_i^{\text{sentiment}}, \boldsymbol{E}_i^{\text{type}}, \boldsymbol{E}_i^{\text{read-time}}, \boldsymbol{E}_i^{\text{scroll}}) \in \mathbb{R}^d \tag{2}$$

Based on the user's historical interactions, denoted by $\mathcal{H}_u$, $\mathcal{R}_u$, and $\mathcal{S}_u$, we derive a sequence of representations that encapsulate the user's long-term interest. Subsequently, we utilize a pointwise

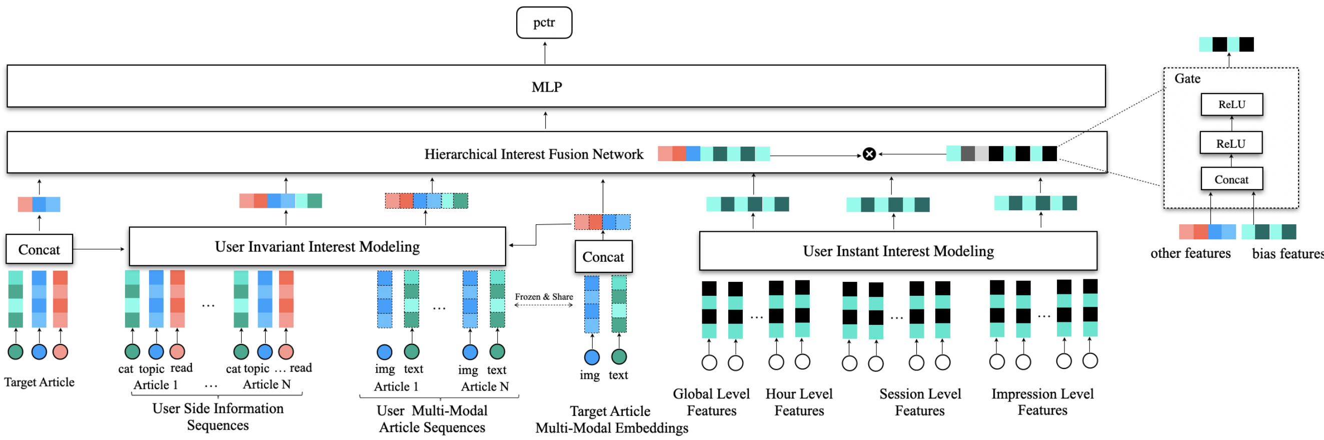

Figure 2: We utilize the multimodal representations along with the side information of items to distill the long-standing interests from user historical behaviors and develop hierarchical features to identity user instant interests. Then, a interest fusion network is proposed to adaptively fuse the hierarchical interests.

aggregation attention mechanism that eliminates the softmax function, traditionally applied to normalize weight sums to one. This modification allows us to retain the detailed nuances of the user's engagement intensity and consumption repetition patterns, providing a more accurate depiction of the user's historical preferences.

$$E_{\mathcal{H}_u} = \{\boldsymbol{e}_1, ..., \boldsymbol{e}_j, ..., \boldsymbol{e}_m\} \in \mathbb{R}^{m \times d} \tag{3}$$

$$\boldsymbol{e}_u = \sum_{\boldsymbol{e}_j \in E_{\mathcal{H}_u}} \text{Attention}(\boldsymbol{e}_j, \boldsymbol{e}_i) \cdot \boldsymbol{e}_j \tag{4}$$

where $\text{Attention}(\boldsymbol{e}_i, \boldsymbol{e}_j) = \text{MLP}(\boldsymbol{e}_j, \boldsymbol{e}_i, \boldsymbol{e}_i - \boldsymbol{e}_j, \boldsymbol{e}_i \odot \boldsymbol{e}_j)$.

Similarly, the multimodal representation of articles revealing the semantic information, is well-suited for capturing the user's long-standing interests. Considering the trade-offs between performance and computational efficiency, we apply PCA (Principal Component Analysis) to reduce the dimension of the image and text representations provided by the official providers. We then concatenate these reduced representations, freeze them and adopt the pointwise aggregation attention mechanism to construct user interest representation.

$$\boldsymbol{e}'_i = \text{Concat}(\boldsymbol{E}_i^{\text{img}}, \boldsymbol{E}_i^{\text{text}}), \boldsymbol{E}'_{\mathcal{H}_u} = \{\boldsymbol{e}'_1, \boldsymbol{e}'_2, ..., \boldsymbol{e}'_m\} \tag{5}$$

$$\boldsymbol{e}'_u = \sum_{\boldsymbol{e}_j' \in E'_{\mathcal{H}_u}} \text{Attention}(\boldsymbol{e}'_j, \boldsymbol{e}'_i) \cdot \boldsymbol{e}'_j \tag{6}$$

Ultimately, the long-standing interest representation of a user can be expressed in the following manner:

$$\boldsymbol{e}_u^{\mathcal{H}} = \text{Concat}(\boldsymbol{e}_u, \boldsymbol{e}'_u) \tag{7}$$

3.2 User Instant Interest Modeling

News evolves dynamically everyday, with the emergence of new and sudden hot topics, which often results in substantial shifts in user interests. Traditional models based on historical behavior sequences struggle to capture these abrupt preferences and encounter cold start problems. A more rational approach is to leverage elaborate feature engineering to uncover instant interests beyond stable interests. To this end, we propose to utilize the hierarchical user behavioral paths to discover the connections and transferability of user interests.

Specifically, we devise several strong features from $(u, q, p, i) \in \mathcal{X}$ in hierarchical manners:

- **Global Non-Personalized Interests**: This represents the intrinsic popularity for articles. It includes features such as the absolute exposure count and its relative changes within the first 48 hours post-publication.
- **Hourly Real-time Non-Personalized Interests**: This captures the popularity of articles at the hourly granularity reflecting the shift in non-personalized interests. It includes features such as hourly absolute exposure count and its relative changes in real-time context.
- **Session-Level Personlized Interests**: Users may engage in multiple impression interactions within a short time span, associated with the same session ID. Behaviors across different impression IDs within a session have certain connections, reflecting the user's consistent interests or regular changes in interests. Therefore, we extract the frequency of the same article's appearance within a session, the time differences between impressions, and compare the differences in the popularity of various articles across impressions.
- **Impression-Level Personlized Interests**: We profile user state and interests through comparative analysis of articles. For instance, by extracting the number of content items a user browses, we can infer their screen size indirectly. Moreover, by contrasting the differences between articles, we reveal the intricate preferences and real-time shifts in user interests.

Further, we propose a unified list-wise feature mining framework to enhance the session and impression level interest representation. Firstly, based on the non-personalized interests, we calculate various popularity metrics for articles, such as exposure times, total read time, total inviews, total pageviews, cumulative hourly inviews, and days since publication. Then, we group by impression and session respectively, sort the articles within each group by their impression time, and calculate the differences in impression time, statistical metrics, and the deviation from the mean values of articles within the group. This approach enables a comprehensive and finely detailed depiction of users' instant interests.

Beyond this, we identify a category of features that can characterize the intrinsic quality of articles, reflecting the editorial preferences of the platform. That is, high-quality content is inclined to be featured in prominent display areas. We convey the characteristics by calculating the average inviews of the article across different impressions.

For use in HUIM, we uniformly discretize these numerical features into 50 bins using equal-frequency binning, and then learn them end-to-end within the model through embeddings. For tree models, they are used directly as inputs for modeling.

$$\boldsymbol{e}_u^{\mathcal{L}} = \text{Concat}(\boldsymbol{E}_{Bin(global)}, \boldsymbol{E}_{Bin(hour)}, \boldsymbol{E}_{Bin(session)}, \boldsymbol{E}_{Bin(impression)}) \tag{8}$$

3.3 Hierarchical Interest Fusion Network

The extracted historical interest representations $\boldsymbol{e}_u^{\mathcal{H}}$ and instant interest representations $\boldsymbol{e}_u^{\mathcal{L}}$ from the impression log express interests in heterogeneous spaces, which require adaptive fusion through the network to achieve accurate user preferences. For this purpose, we propose an adaptive interest fusion network. Firstly, interest representations along with contextual information, are normalized using batch normalization.

$$\boldsymbol{e}_{\text{context}} = \text{BatchNormalization}(\text{Concat}(\boldsymbol{e}_u^{\mathcal{H}}, \boldsymbol{e}_u^{\mathcal{L}}, \boldsymbol{e}_c)) \tag{9}$$

In practice, we experiment with two fusion method. The first method involves passing the $\boldsymbol{e}_{\text{context}}$ directly through MLP to achieve the fused representation.

$$\boldsymbol{e}_{\text{output}} = \text{MLP}(\boldsymbol{e}_{\text{context}}) \tag{10}$$

The second method adpots the attention mechanism to dynamically capture the user's instance-wise interests where Gate is two-layer MLPs with ReLU.

$$\boldsymbol{e}_{\text{output}} = \text{MLP}(\text{Gate}(\boldsymbol{e}_{\text{context}}) \odot \text{Concat}(\boldsymbol{e}_u^{\mathcal{H}}, \boldsymbol{e}_u^{\mathcal{L}}, \boldsymbol{e}_c)) \tag{11}$$

The second method tends to perform better in practice, as it can capture preference differences at the sample granularity in addition to evaluating the feature importance. Finally, we use the MLP to output the logits.

$$\hat{r} = \text{MLP}(\boldsymbol{e}_{\text{output}}) \tag{12}$$

3.4 Learning Objective

To better model user interests, we refine the pointwise log loss by experimenting with pairwise loss, listwise loss, and query-level cross entropy loss following [2]. Among these, the query-level cross entropy loss performed the best, effectively balancing calibration and ranking capabilities.

Since the internal structure of the data is known, it can be assumed that the predictions in various groups are different. This can be modeled by adding a shift group to each formula prediction for a group:

$$\hat{p} = \sigma(\hat{r} + \delta_{\text{imp}}) \tag{13}$$

Then, the group-aware logloss is formulated as follows:

$$LogLoss_{imp} = -\sum_{\text{imp}} \Big(\sum_{\text{article in imp}} y_i \log(\hat{p}_i) + (1 - y_i) \cdot \log(1 - \hat{p}_i) \Big) \tag{14}$$

The δ_{imp} parameter is jointly optimized for each group during the training.

The ultimate objective loss function is a weighted sum of the global log loss and the group log loss. α is set to 0.95 by default.

$$QueryCrossEntropy(\alpha) = (1 - \alpha) \cdot LogLoss + \alpha \cdot Logloss_{\text{imp}} \tag{15}$$

Finally, we achieved a score of 0.873 for the single HUIM model, 0.8815 for the blending model of HUIM and catboost and ranked 2nd place on the final leaderboard.

4 EVALUATION

We employ the official partitioning of the training and validation datasets illustrated in Figure 1 for our offline experiments. The improvements observed in offline metrics are in close alignment with the enhancements seen online. To enhance iteration efficiency due to the vastness of the original data, we randomly sample 5% of the user IDs from training and validation datasets respectively, preserving the entirety of the data for these users. This strategy ensures that the offline gains are consistently reflective of the online outcomes.

For the historical behaviors utilized in the HUIM, we augment the history for both the validation and test sets. Specifically, when constructing the validation set's history, in addition to the official provided validation history, we augment it with the training history and behaviors. The test set follows the same principle. We set the length of historical behavior to 200 and set the PCA reduction dim to 64, as increasing it further did not yield significant improvements.

The HUIM model is trained with the AdamW optimizer, with a batch size of 10,240, a learning rate of 0.001, and an epoch of 2. Early stopping is applied to avoid overfitting.

To handle the complete dataset, we leverage a system equipped with 1 TB of memory and 8 NVIDIA A100 GPUs, thereby optimizing our data processing and model training speeds. For iterating on our HUIM, we relied on PyTorch's FuxiCTR [8] framework, whereas for tree-based models, we adopt the CatBoost [2] algorithm.

4.1 Results

We conduct our training on the training set and evaluate the model's performance on the validation set during the offline iteration. For online submission, we employ two strategies to integrate the data from the 2-fold training and validation sets.

- We train two models: one on the training set with validation for early stopping, and another on the validation set with training for early stopping. Their predictions on the test set are averaged. This method is practical for neural networks.
- We merge the training and validation sets to train fully on a pre-set number of iteration steps. This approach is practical for tree-based models.

Table 1: Results of our models

Method	Online G-AUC
HUIM with LogLoss	0.873
Catboost1 with LogLoss	0.861
Catboost2 with PairWiseRankLoss	0.873
Catboost3 with Query-Level Loss	0.879
Blending (HUIM & Catboost 2/3)	0.881

Table 2: The top 6 teams on the final leaderboard

Rank	Team	Score
1	:D	0.8924
2	BlackPearl	0.8815
3	Tom3TK	0.8707
4	axi2017	0.8692
5	hrec	0.8667
6	FeatureSalad	0.8513

By employing the two-fold method, we are able to further enhance online metrics by nearly 1%.

The detailed performance of HUIM and catboost are shown in Table 1. The HUIM model utilizes all of CatBoost's features in addition to historical behavior sequences. By comparing the performance of HUIM and CatBoost both using logloss, the superior performance of HUIM indicates the effectiveness of historical interests. For CatBoost models using logloss, pairwise and query-level loss, it is evident that contrasting the difference within impressions in the loss can enhance the model's ranking performance.

Additionally, the integration of HUIM and CatBoost initially showed significant benefits, especially when instant interest features were less developed, with improvements exceeding 1%. However, as these features became more comprehensive, the benefits of integration diminished. This suggests that in cold start situations, capturing short-term interests is more crucial than historical modeling.

Lastly, due to tight schedule and the fact that CatBoost has already incorporated query-level loss, we did not use query-level loss in HUIM. It is anticipated that integrating query loss into HUIM could potentially enhance performance beyond that of the best single CatBoost model, which can be reserved for future research.

The final rankings of the top 6 teams are shown in Table 2.

4.2 Ablation Study

We conduct the ablation study in the 5 percent datasets for the HUIM in Table 3. For the ablation for historical interest modeling, we can observe that pre-trained multimodal representations and auxiliary information are crucial for capturing users' historical preferences. In addition, the randomly initialized user id and article id information are virtually useless in cold-start scenarios.

When comparing the historical interest modeling with instant interest modeling, we find that instant interests are more important, especially the fine-grained article comparison at the session and impression level.

Table 3: Ablation study of key components for HUIM, Top: ablation for historical interest modeling. Middle: ablation for instant interest modeling. Bottom: ablation for interest fusion network.

Component	G-AUC Decrease
w/o Multimodel Sequence	-1.26%
w/o side information Sequence	-0.6%
w/o article id and user id	-0.04%
w/o Historical Interest Modeling	-1.45%
w/o Global Features	-0.7%
w/o Hourly Real-time Features	-0.5%
w/o Session-Level Features	-0.8%
w/o Impression-Level Features	-3.5%
w/o List-wise Features	-3.2%
w/o Interest Fusion Network	-0.6%

Lastly, it should be noted that the data provided contains a certain degree of future information leakage, evident in the cumulative view and click features of the articles (e.g., total inviews) provided by the official and the impression details in the test logs (e.g., 48-hour inviews and co-occurrence of articles at session-level logs). By completely excluding these features, our single model achieves a score of 0.822.

5 CONCLUSION

In this paper, we propose a hierarchical interest model leveraging both historical invariant interests and short-term rapidly changing interests for the 2024 ACM RecSys Challenge. Our team, BlackPearl, achieved a score of 0.8815 and ranked 2nd place on the final leaderboard.

REFERENCES

[1] Jianxin Chang, Chenbin Zhang, Yiqun Hui, Dewei Leng, Yanan Niu, Yang Song, and Kun Gai. 2023. Pepnet: Parameter and embedding personalized network for infusing with personalized prior information. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 3795–3804.

[2] Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. 2018. CatBoost: unbiased boosting with categorical features. *Advances in neural information processing systems* 31 (2018).

[3] Maksims Volkovs, Guangwei Yu, and Tomi Poutanen. 2017. Dropoutnet: Addressing cold start in recommender systems. *Advances in neural information processing systems* 30 (2017).

[4] Zhiqiang Wang, Qingyun She, and Junlin Zhang. 2021. Masknet: Introducing feature-wise multiplication to CTR ranking models by instance-guided mask. *arXiv preprint arXiv:2102.07619* (2021).

[5] Xue Xia, Pong Eksombatchai, Nikil Pancha, Dhruvil Deven Badani, Po-Wei Wang, Neng Gu, Saurabh Vishwas Joshi, Nazanin Farahpour, Zhiyuan Zhang, and Andrew Zhai. 2023. Transact: Transformer-based realtime user action model for recommendation at pinterest. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 5249–5259.

[6] Guorui Zhou, Na Mou, Ying Fan, Qi Pi, Weijie Bian, Chang Zhou, Xiaoqiang Zhu, and Kun Gai. 2019. Deep interest evolution network for click-through rate prediction. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 5941–5948.

[7] Guorui Zhou, Xiaoqiang Zhu, Chenru Song, Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li, and Kun Gai. 2018. Deep interest network for click-through rate prediction. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 1059–1068.

[8] Jieming Zhu, Jinyang Liu, Shuai Yang, Qi Zhang, and Xiuqiang He. 2021. Open benchmarking for click-through rate prediction. In *Proceedings of the 30th ACM international conference on information & knowledge management*. 2759–2769.