TDFM and TAFM: Time-Aware and Feature Fusion-Based Deep Recommendation Models for Short Videos

Bing Li[1,2,3(✉)], Yuqi Hou[1], and Biao Yang[4]

[1] School of Software, Jiangxi Normal University, Nanchang 330022, Jiangxi, China
004970@jxnu.edu.cn
[2] Research Centre for Management Science and Engineering, Jiangxi Normal University, Nanchang 330022, Jiangxi, China
[3] Jiangxi Provincial Engineering Research Center of Blockchain Data Security and Governance, Nanchang 330022, Jiangxi, China
[4] School of Digital Industries, Jiangxi Normal University, Nanchang 330022, Jiangxi, China

Abstract. With the increasing growth of technological innovation and the emergence of technological achievements, 5G technology is driving the transformation of social and information dissemination from text to video, and video information, live streaming and short video will usher in new development opportunities. With the continuous development of the two head platforms TikTok and Kwai, the short video market pattern is gradually stabilizing, the user coverage rate is increasing and the growth rate is starting to slow down. How to recommend short videos that are more in line with users' preferences has become a topic of more concern for platforms and users nowadays. However, the existing short video recommendation model is only based on user and item features, and lacks the perception of time features, resulting in the existing short video recommendation model is not ideal and needs to be improved. A time-aware and feature fusion-based deep recommendation model for short videos, including the TDFM model and TAFM model, is proposed in the paper. The two proposed methods are compared iteratively with traditional recommendation algorithms. This paper proposes a better method with better results. We make our experimental code public so that our experiments can be verified and reproduced. (https://github.com/lbxd123/Time-aware-recommendation.git)

Keywords: Time-Aware · Feature fusion · Short Video · Recommendation

1 Introduction

In the era of information explosion, recommender systems have become a prominent force, providing personalized and accurate information services through in-depth analysis of user behavior and interests, and successfully applied in social

Y. Cao and X. Shao (Eds.): DIONE 2023, LNICST 515, pp. 263–275, 2025.
https://doi.org/10.1007/978-3-031-80713-8_20

networks, e-commerce, music and movies, news and other fields. Well-known platforms such as Taobao, Facebook, Amazon, Spotify, etc. have optimized user experience, improved user satisfaction and platform profitability through recommender systems, and at the same time promoted the innovation and development of related industries, establishing themselves as pioneers in the information age. As shown in Fig. 1, the impact of recommender systems spans multiple domains.

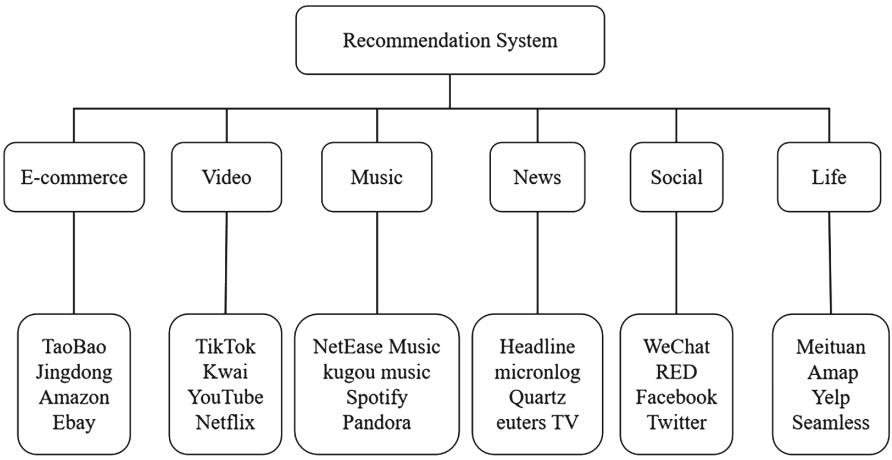


Fig. 1. The main application areas.

Short videos have become a popular way of content distribution. Although the short video market has stabilized due to the continued growth of platforms such as TikTok and Kwai, the rate of growth has begun to slow down as the reach of users expands. As a result, recommending short videos that match user preferences has become an important focus for both platforms and users. Taking YouTube as an example, the platform leverages user interaction data for Top-N video recommendations [1] and employs a two-layer deep neural network, including recall and ranking, to swiftly optimize and filter candidate videos, thereby enhancing the efficiency of the recommendation system [2].

In recent years, there has been significant progress in recommender system research, but the current service recommendation models still lack depth in considering time-aware factors. They predominantly rely on user behavior characteristics to predict the service items B that user A might need, overlooking time requirements. Scholars have proposed time-aware service recommendation methods, such as Ngaffo (2020), Tong(2021), Li (2021), Alabduljabbar, R (2023), Wen (2022) [3–7]. However, existing research mainly focuses on the time factors in traditional service recommendations, neglecting the "user-item-time" matching relationship and calling for a more comprehensive time-based service recommendation approach. Furthermore, to enhance model performance, scholars such as

Hu B(2020), Liang B(2023), Yu R (2021), Chen M(2021) have proposed multi-feature fusion methods, significantly improving the performance of the models [8–12].

This study proposes a new method for personalized recommendation of short videos using deep learning techniques. The method includes two aspects: time awareness and feature fusion, aiming to improve the accuracy and effectiveness of recommendations. By fusing user's historical interaction data and time information, utilizing deep neural networks for feature learning and interaction modeling, two end-to-end deep learning models, TDFM and TAFM, are formed. Experimental results show that compared with models without time-related factors, our model significantly improves the precision, recall rate, and other evaluation metrics of recommendations.

2 Background Knowledge

2.1 Recommendation Algorithm

Since users' preferences are dynamic, it is difficult to express them in words. Recommendation algorithms can help us better understand users' preferences and accurately recommend items to them. Collaborative filtering based on user's historical behavioral interaction data is one of the mainstream recommendation algorithms [13], mainly divided into user CF, item CF and model-based recommendation. Model-based recommendation algorithms train models based on user behavior data, and then calculate the recommendation results for users by the models. Commonly used training models include association algorithm [14], matrix decomposition algorithm [15], graph model [16] and implicit semantic model [17].

2.2 DeepFM

The DeepFM model is a recommendation system that integrates Factorization Machines (FM) and Deep Neural Networks (DNN) [18]. In its model diagram, the input layer encompasses various raw features of users and items. The first-order features are obtained through a linear model, while the FM component learns second-order cross features. The embedding layer embeds raw features into low-dimensional dense vectors, and the cross network achieves high-order feature interactions through a multi-layer neural network. The fully connected layer is employed for non-linear mapping, ultimately producing recommendation scores in the output layer. This design aims to comprehensively utilize both linear and non-linear relationships, providing a holistic representation of the complex feature interactions between users and items to enhance the performance of the recommendation system. For detailed steps, refer to Fig. 2.

DeepFM stands out with its simple structure and ease of operation, supporting end-to-end learning without the need for additional artificial features. Through the joint training of these two components, DeepFM can fully leverage sparse feature data, thereby improving prediction accuracy and the model's

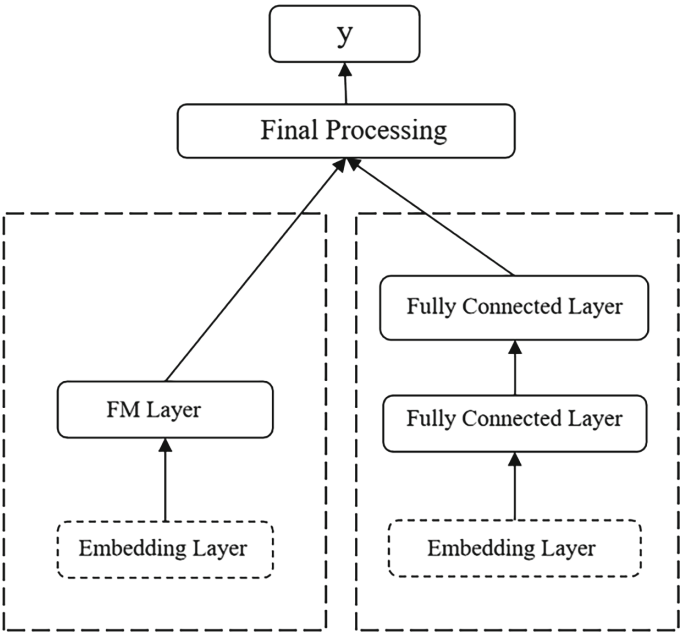


Fig. 2. DeepFM Model Structure.

generalization capability. However, with the linear growth of feature numbers, the model often introduces a significant amount of redundant computation.

2.3 AFM

The attention mechanism, based on the natural tendency of humans to selectively focus on certain information, has been widely used in various deep learning fields [19]. It can help models allocate different weights to input features, extract key information, and improve prediction performance. In 2017, Zhejiang University and National University of Singapore jointly proposed an AFM model with attention mechanism as an improved version of the FM model. The AFM model introduces the attention mechanism into the "feature interaction layer" and "output layer", dynamically learns the importance of each feature, thus improving overall performance, and achieves better expression capability and interpretability through adaptive learning of feature importance. The AFM model can be used for classification, regression, and ranking problems, and through the attention network mechanism, it optimizes FM, improves feature expression ability, and makes it more interpretable.

3 Deep Service Recommendation Model Construction

The deep recommendation model is a method that uses deep learning technology for personalized recommendation of short videos [20]. It combines time aware-

ness and feature fusion to enhance understanding of user interests and improve the effectiveness of short video recommendations. The model consists of several main parts: feature extraction, time-aware modeling, feature fusion, and end-to-end learning for video recommendations. Overall, this model combines various techniques to better understand user interests and improve the accuracy and coverage of short video recommendations.

3.1 Temporal Characteristics of User Behavior

The importance of time in influencing user preferences is increasingly recognized in recommender systems. Traditional recommender systems often overlook the temporal dimension and focus solely on user-item association, while deep service recommendations explore underlying temporal patterns [21]. The figure provides a formal representation of the relationship between users, items, and active time (Fig. 3).

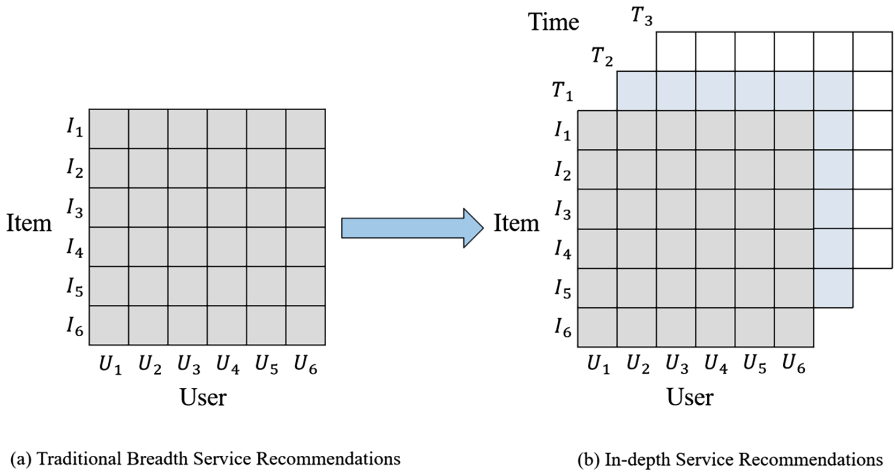


Fig. 3. User-Item-Time 3D database.

To obtain a high-quality low-dimensional feature representation of users, services, and time in the recommendation system, semantic transformation of timestamps and analysis of the user's time tensor are performed [22]. The data is mapped into three-dimensional space to construct a user-item-time database. Statistical analysis of the database provides insight into user demand cycles, frequency, and preferences over time. Time series integration is done through a non-linear approach and incorporates the time factor into the recommendation system. Long-term user preference modeling based on dynamic features is explored, along with in-depth learning of changing demand characteristics for service quality and level.

3.2 TDFM Deep Service Recommendation Model

The TDFM model is a combination of DeepFM and time, aimed at incorporating temporal dynamics into the recommendation process. The features of user behavior data in the dataset are encoded after undergoing pre-processing. Interactive information from the user's historical actions is used to extract relevant features. A user-item-time feature matrix is constructed, representing the interactions between users, items, and time. Three feature vectors, representing user, item, and time, are randomly initialized. These vectors are then fused and vertically stitched together to form a matrix, combining the three dimensions. Next, the spliced matrices are passed through MLP and FM aggregation layers for correlation prediction, respectively, and then the obtained predicted values are spliced. The performance of a TDFM is assessed by means of assessment indicators. These metrics assess the accuracy, relevance, or other aspects of the model's recommendations. The model structure is illustrated in the figure below (Fig. 4):

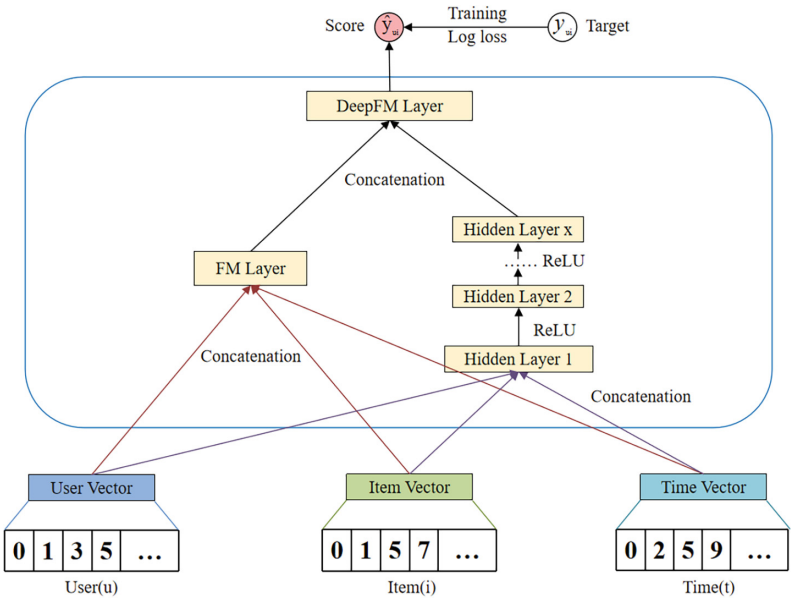


Fig. 4. TDFM deep learning model.

Sparse Features: Because the category features do not directly feed to the model, the user, item, and time features are first encoded and processed. Here hard index coding is performed to obtain the vector feature matrix, and this layer represents all category features for coding processing.

Dense Embedding: A dense embedding layer is used to convert the high-dimensional dense feature matrix into a lower-dimensional dense vector. Each

dense vector is vertically stitched together and serves as input for the Deep and FM layers.

FM Layer: The FM layer takes the primitive feature vector as input. It performs feature interactions by multiplying the features by two and assigning weights to the resulting interactions. This layer allows the model to learn the low-order feature effects.

Hidden Layer: The dense vectors from the Deep part are vertically stitched and processed through a hidden layer. This layer consists of multi-layer linear mappings and non-linear transformations. The output of the hidden layer is typically mapped to one dimension, as it needs to be combined with the results from the FM layer.

Output Units: The output layer combines the results from the FM layer and the hidden layer. It fuses the low-order and high-order features to capture both their effects. A sigmoid non-linear transformation is applied to the combined output, resulting in a predicted probabilistic output.

3.3 TAFM Deep Service Recommendation Model

TAFM is the product of combining AFM with time. Based on FM, a weight is assigned to the result of FM, which is obtained by learning and available for represent the attention paid to different features between crossovers. The input and embedding layers are identical to TAFM, where the input features are hard-indexed and encoded, and the feature matrix is embedded into the dense vector. The whole model framework is shown in the following figure (Fig. 5).

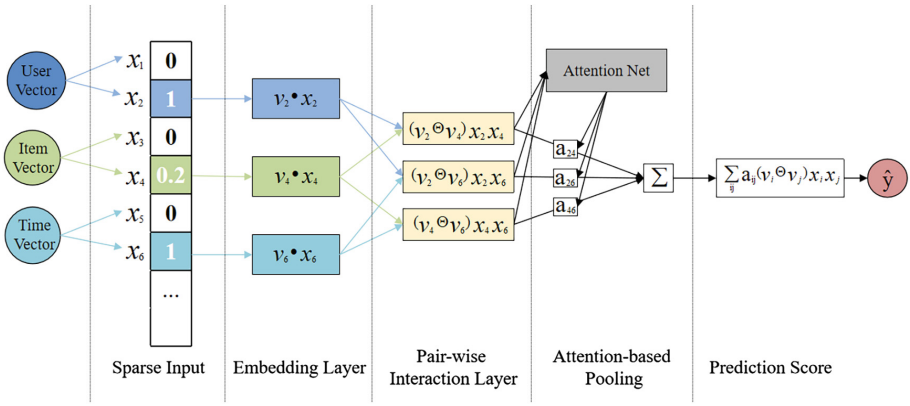


Fig. 5. TDFM deep learning model.

$$\widehat{y} = \delta\left(w0 + \sum_{i=1}^{n} w1_i x_i + P^T \sum_{i=1}^{n} \sum_{j=i+1}^{n} a_{ij}(v_i \odot v_j) x_i x_j\right) \tag{1}$$

Pair-wise Interaction Layer: The main idea is to model the combinatorial features by taking all the input embedding vectors m and interacting them two by two to output $m(m-1)/2$ vectors, which have the same dimensionality as their vectors. It is equivalent to the computational logic of FM in a neural network, where ξ denotes the output from the embedding layer and $\odot$ denotes the inner product with two vectors.

$$f_{PI}(\xi) = \{(v_i \odot v_j)x_i x_j\}_{(i,j)\in R_x} \tag{2}$$

Attention-based Pooling: This pooling method is based on an attention mechanism that adaptively weights the input data to enhance the models' performance and accuracy. AFM implements the Attention mechanism by adding a weighted sum after the Interacted vector. a_{ij} is the attention between feature i and feature j, which indicates how different combined features contribute to the prediction of the degree of contribution of different combined features. The form is as follows:

$$f_{Att}(f_{PI}(\xi)) = \sum_{(i,j)\in R_x} a_{ij}(v_i \odot v_j)x_i x_j \tag{3}$$

A multiplicative attention mechanism is introduced, where weight is obtained after normalization through a softmax layer. w and b can be considered as weights and bias terms of a linear layer whose input is the hidden vector length k. The output is a hyperparameter, assumed to be t, so that $W \in R^{t\times k}$, $b \in R^t$, and $h \in R^t$. Where x_i x_j is a user-item-time three-dimensional feature vector stitched together.

$$a'_{ij} = h^T ReLU(W(v_i \odot v_j)x_i x_j + b) \tag{4}$$

$$a_{ij} = Softmax(a'_{ij}) = \frac{exp(a'_{ij})}{\sum_{(i,j)\in R_x} exp(a'_{ij})} \tag{5}$$

4 Experiment

In this chapter, compares the depth service model we propose with the original model, and the experimental results show that the model is effective.

4.1 Dataset

The validity of the model was evaluated using the publicly available dataset KuaiRec. The KuaiRec dataset is a real dataset created by CSU and the Racer team, consisting of recommendation logs from the video-sharing mobile app Racer. It spans from July 5 to September 5, 2020, with a data consistency rate of 99.6% [23]. The dataset contains a large volume of user-item interactions and includes side information such as item categories and social network information. User-video features and interaction duration are also extracted and

encoded to provide further insights into user preferences. The dataset is valuable for evaluating and improving recommendation systems, especially in the context of video-sharing platforms (Table 1).

Table 1. Dataset

Dataset	Users	Items	Interactions	Density
KuaiRec	1411	3327	4676570	99.6%

4.2 Evaluation Indicators

The paper uses four popular evaluation metrics to assess model performance and accuracy: average loss value, precision, recall, and accuracy. These metrics allow for a comprehensive assessment of different aspects of model performance, enabling researchers to compare and select models based on their performance on these metrics (Table 2).

Table 2. Basic concepts

	P(Positive)	N(Negative)
T(True)	TP	FN
F(False)	FP	TN

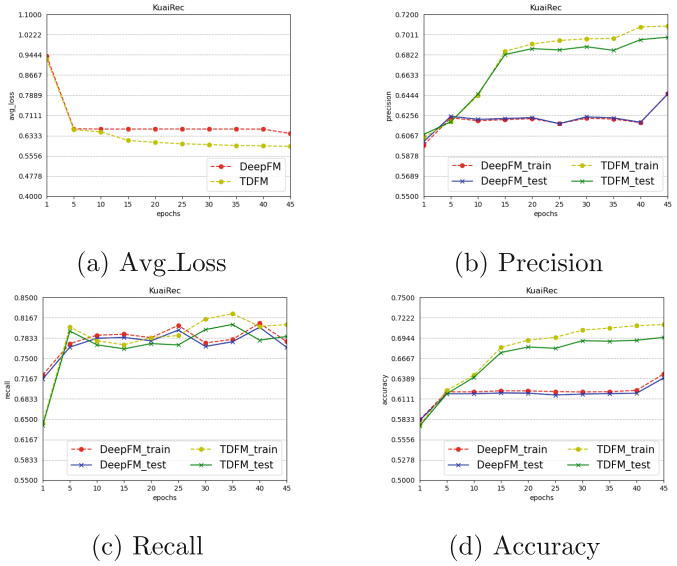
The definitions of the three indicators are shown below:

$$Precision = \frac{TP}{TP + FP} \quad Recall = \frac{TP}{TP + FN} \quad Accuracy = \frac{TP + TN}{P + N} \tag{6}$$

4.3 Performance Comparisons

The effectiveness of TDFM and TAFM models is validated by evaluating them on the KuaiRec dataset, a popular dataset in the short video domain. The performance of the improved models is compared with that of the original models (DeepFM and AFM) using the four evaluation metrics. Results show that the TDFM and TAFM models outperform the original models, demonstrating their effectiveness across different recommendation domains. An experimental comparison chart is presented to support this finding (Figs. 6 and 7).

The experimental results demonstrate that the TDFM and TAFM models outperform their original versions across all metrics, including accuracy, recall, average loss, and precision. The improvements in accuracy and recall suggest



(a) Avg_Loss (b) Precision

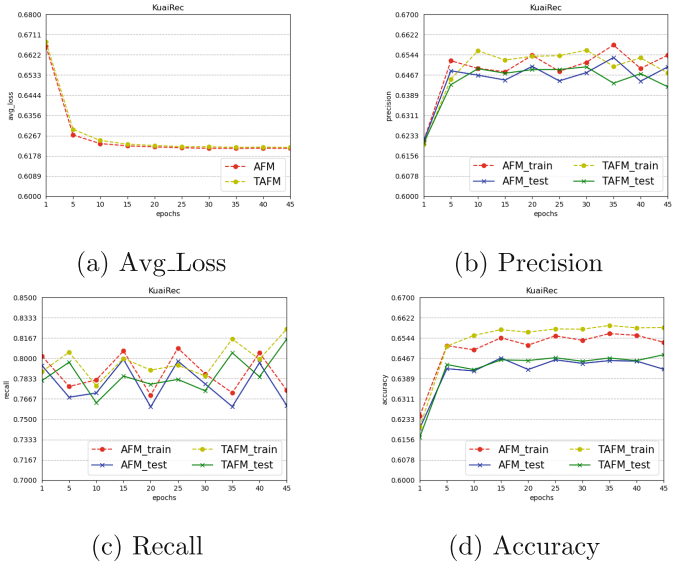
(c) Recall (d) Accuracy

Fig. 6. DeepFM and TDFM models.

that the new models more effectively capture user interests, thereby enhancing the quality of recommendations. Simultaneously, the reduction in average loss reflects better training performance of the models, while the enhancement in precision indicates more accurate predictions of user preferences. Visualized through experimental comparison charts, the superiority of TDFM and TAFM models over DeepFM and AFM models is evident. These findings affirm the effectiveness of introducing time factors in enhancing the performance of recommender system models and provide robust support for the applicability of the models across diverse recommendation domains (Table 3).

Table 3. The mean of results under KuaiRec dataset experiments.

	TDFM		TAFM	
	Train	Test	Train	Test
Precision	0.6797	0.6741	0.6538	0.6475
Recall	0.7967	0.7844	0.7873	0.7753
Accuracy	0.6832	0.6720	0.6557	0.6435



(a) Avg_Loss (b) Precision

(c) Recall (d) Accuracy

Fig. 7. AFM and TAFM models.

5 Conclusion

In the paper, the TDFM and TAFM models are proposed as improvements over the traditional DeepFM and AFM models. These proposed models introduce a temporal dimension by incorporating user-item-time 3D feature vector fusion methods, which aim to enhance the accuracy of the models compared to the traditional user-item 2D feature fusion.

By conducting experiments on both the KuaiRec dataset, the proposed improvements are empirically validated and shown to be more effective in the context of recommendation systems. The results of the experiments demonstrate that the TDFM and TAFM models outperform their respective traditional counterparts (DeepFM and AFM) in terms of the evaluation metrics.

Building upon these findings, the paper plans to further propose time-aware models for service recommendations. The objective of these future models is to enhance both the precision and timeliness of service recommendations. By considering the temporal aspect in the recommendation process, the proposed models aim to offer users with more precise and current recommendations.

Funding. This work was funded by several organizations, including the Jiangxi Provincial Natural Science Foundation (20224BAB202023), the National Natural Science Foundation of China (72161020), the Jiangxi Social Science Foundation Project (21GL44), the Science and Technology Research Project of Jiangxi Provincial Education Department (GJJ2200333), and the Youth Fund Project of Humanities and Social Sciences in Colleges and universities of Jiangxi Province (GL19223).

References

1. Covington, P., Adams, J., Sargin, E.: In Deep neural networks for YouTube recommendations. In: ACM Conference on Recommender Systems, pp. 191–198 (2016)
2. Davidson, J., et al.: The YouTube video recommendation system. In: Proceedings of the Fourth ACM Conference On Recommender Systems, Association for Computing Machinery: Barcelona, Spain, pp. 293–296 (2010)
3. Noulapeu Ngaffo, A., El Ayeb, W., Choukair, Z.: A time-aware service recommendation based on implicit trust relationships and enhanced user similarities. J. Ambient Intell. Human. Comput. **12**(2), 3017–3035 (2021). https://doi.org/10.1007/s12652-020-02462-5
4. Tong, E., Niu, W., Liu, J.: A missing QoS prediction approach via time-aware collaborative filtering. IEEE Trans. Serv. Comput. **15**(6), 3115–3128 (2021)
5. Li, H., Han, D.: A novel time-aware hybrid recommendation scheme combining user feedback and collaborative filtering. IEEE Syst. J. **15**(4), 5301–5312 (2021)
6. Alabduljabbar, R., Alshareef, M., Alshareef, N.: Time-aware recommender systems: a comprehensive survey and quantitative assessment of literature. IEEE Access **11**, 45586–45604 (2023)
7. Wen, W., Liang, F.: Deep structured state learning for next-period recommendation. IEEE Trans. Neural Netw. Learn. Syst. **35**(1), 680–692 (2022)
8. Hu, B., Gao, B., Woo, W.L., et al.: A lightweight spatial and temporal multi-feature fusion network for defect detection. IEEE Trans. Image Process. **30**, 472–486 (2020)
9. Liang, B., Meng, X., Zhang, Y.: Exploring time-aware multi-pattern group venue recommendation in LBSNs. ACM Trans. Inform. Syst. **41**(3), 1–31 (2023)
10. Yu, R.Y., et al.: Cffnn: Cross feature fusion neural network for collaborative filtering. IEEE Trans. Knowl. Data Eng. **34**, 4650–4662 (2022)
11. He, L., Chen, H., Wang, D., Jameel, S., Yu, P., Xu, G..: Click-through rate prediction with multi-modal hypergraphs. In: Proceedings of the 30th ACM International Conference on Information & Knowledge Management, Association for Computing Machinery: Virtual Event, Queensland, Australia, pp. 690–699 (2021)
12. Chen, M., Li, Y., Zhou, X.: Conet: co-occurrence neural networks for recommendation. Future Gener. Comp. Sys. **124** 308–314 (2021)
13. He, X., Liao, L., Zhang, H., Nie, L., Hu, X., Chua, T.S.: Neural collaborative filtering. In: Proceedings of the 26th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee: Perth, Australia, pp. 173–182 (2017)
14. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules in large databases. In: Proceedings of the 20th International Conference on Very Large Data Bases, Morgan Kaufmann Publishers Inc, 487–499 (1994)
15. Koren, Y.: Factorization meets the neighborhood: a multifaceted collaborative filtering model. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery: Las Vegas, Nevada, USA, pp. 426–434 (2008)
16. Guo, Q.Y., et al.: A survey on knowledge graph-based recommender systems. IEEE Trans. Knowl. Data Eng. **34**(8), 3549–3568 (2022)
17. Wu, L., He, X., Wang, X., et al.: A survey on accuracy-oriented neural recommendation: From collaborative filtering to information-rich recommendation[J]. IEEE Trans. Knowl. Data Eng. **35**(5), 4425–4445 (2022)

18. Guo, H., Tang, R., Ye, Y., Li, Z., He, X.: Deepfm: a factorization-machine based neural network for ctr prediction. In: Proceedings of the 26th International Joint Conference on Artificial Intelligence, AAAI Press: Melbourne, Australia, pp. 1725–1731 (2017)
19. Xiao, J., Ye, H., He, X., Zhang, H., Wu, F., Chua, T.-S.: Attentional factorization machines: learning the weight of feature interactions via attention networks. In Proceedings of the 26th International Joint Conference on Artificial Intelligence, AAAI Press: Melbourne, Australia, pp. 3119–3125 (2017)
20. Zhang, S., Yao, L., Sun, A., et al.: Deep learning based recommender system: a survey and new perspectives[J]. ACM Comput. Surv. (CSUR) **52**(1), 1–38 (2019)
21. Mezni, H., Ait Arab, S., Benslimane, D., Benouaret, K.: An evolutionary clustering approach based on temporal aspects for context-aware service recommendation. J. Ambient. Intell. Humaniz. Comput. **11**(1), 119–138 (2020)
22. Zhang, Z., Pan, X., Dong, H., et al.: Behavior modeling network with inter-layer attention for human mobility prediction. In: 2023 IEEE 6th International Conference on Electronic Information and Communication Technology (ICEICT). IEEE, pp. 186–191 (2023)
23. Gao, C., et al.: Kuairec: A fully-observed dataset and insights for evaluating recommender systems. In: Proceedings of the 31st ACM International Conference on Information and Knowledge Management, Association for Computing Machinery: Atlanta, GA, USA, pp. 540–550 (2022)