

## Research Article

# XGBDeepFM for CTR Predictions in Mobile Advertising Benefits from Ad Context

Han An  and Jifan Ren 

*School of Economics and Management, Harbin Institute of Technology, Shenzhen 518055, China*

Correspondence should be addressed to Jifan Ren; renjifan@hit.edu.cn

Received 28 February 2020; Revised 2 April 2020; Accepted 11 April 2020; Published 27 April 2020

Guest Editor: Shianghu Wu

Copyright © 2020 Han An and Jifan Ren. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The problem of click-through rate (CTR) prediction in mobile advertising is one of the most informative metrics used in mobile business activities, such as profit evaluation and resource management. In mobile advertising, CTR prediction is essential but challenging due to data sparsity. Moreover, existing methods often have difficulty in capturing the different orders of feature interactions simultaneously. In this study, a method was developed to obtain accurate CTR prediction by incorporating contextual features and feature interactions. We initially use extreme gradient boosting (XGBoost) as a feature engineering phase to select highly significant features. The selected features are mobile contextual attributes including time contextual, geography contextual, and other contextual attributes (e.g., weather condition) in actual mobile advertising situations. Our model, XGBoost deep factorization machine- (FM-) supported neural network (XGBDeepFM), combines the power of XGBoost for feature selection, FM for two-order cross feature interaction, and the deep neural network for high-order feature learning in a united architecture. In a mobile advertising condition, our methods lead to significantly accurate CTR prediction in “wide and deep” type of model. In comparison with existing models, many experiments on commercial datasets show that the XGBDeepFM model has better value of area under curve and improves the effectiveness and efficiency of CTR prediction for mobile advertising.

## 1. Introduction

The task of click-through rate (CTR) prediction is crucial in advertising and recommendation areas; its main goal is to maximize the clicks to improve advertising revenue or user satisfaction [1–3]. In advertising area, CTR is an important indicator for measuring the effectiveness of advertising displays [4]. Advertiser’s revenue relies heavily on the capability of CTR prediction. In recommendation area, the recommended items returned to users can be ranked by the predicted CTR [5]. This predicted probability helps recommendation systems know the users’ interest on specific items such as news [6, 7], movies [8], tags [9], or commercial items [10], which influence the subsequent decision-making [10]. Recommendation solutions can be classified in terms of collaborative, content-based, knowledge-based, demographic, and hybrid [11]. Each strategy can benefit from the task of CTR prediction [12].

One of the core problems that mobile advertising strives to solve is providing the right ads to the right people at the right time and in the right context. Users’ attention time has been greatly reduced; thus, no one has the time to watch useless and intrusive advertisements. Nevertheless, the answer may be in the hands of marketers, especially in a dynamic mobile world. Mobile contextual advertising is not only about finding the right users in the right context, including time, geography, and weather, but is also about connecting the advertisement with the user in ad context and providing a pleasant experience. Accurate CTR prediction is vital to marketers based on contextual features. Another key challenge for CTR prediction is learning low- and high-order feature interactions behind user behavior in a certain context. Some feature interactions are easy to capture. Low-order feature interactions (less than two orders) can be designed by experts’ prior experience. However, high-order feature interactions can be difficult to understand. These

deep-feature interactions can be learned by deep neural networks (DNNs).

At present, researchers have proposed different methods of CTR prediction. Some ideas attempt to solve the two-order feature interactions. For example, a logistic regression (LR) model has been used to predict the CTR on Google Ads [13]. Factorization machines (FMs) have been used to consider two-order feature interactions [14]. In recent years, DNNs have been popular because of their capability to learn high-order feature interactions. For example, Zhang et al. studied feature representations and proposed the FM-supported neural network (FNN) [15]. Qu et al. proposed the product-based neural network (PNN), which learns high-order feature interactions by introducing a product layer [16]. Cheng et al. combined “wide” and “deep” (W&D) components in a W&D model for low- and high-order feature interactions [2].

The remainder of this paper is organized as follows. In Section 2, the extreme gradient boosting (XGBoost) deep FM-supported neural network (XGBDeepFM) is proposed, considering the problems in contemporary research and the characteristics in the mobile advertising dataset. In Section 3, the experiment is designed, a comparison experiment is performed for CTR prediction, and the effectiveness and efficiency of the XGBDeepFM model are analyzed. Finally, Section 4 concludes the study.

## 2. Materials and Methods

According to the core challenge of computational advertising proposed by Border [3], the best match between a given user in a given context and a suitable advertisement should be determined.  $((a_i, u_i, c_i), y_i)$  denotes the instance  $i$  of the dataset, where  $a_i$  denotes the advertisement,  $u_i$  represents the user,  $c_i$  denotes the context, and  $y_i \in \{0, 1\}$  is the label of the clicking label.

We propose a united approach, namely, the XGBDeepFM model, which benefits from prior information from context and high-order feature interactions, as shown in Figure 1.

Our approach consists of three components, namely, the XGBoost, FM, and deep components. By using these three components, the proposed XGBDeepFM model can realize the full interactive combination modeling of the bilateral features (i.e., ad and user features) and contextual features.

$$\hat{y} = \text{sigmoid}(y_{\text{XGBoost}} + y_{\text{FM}} + y_{\text{DNN}}). \quad (1)$$

$\hat{y}$  denotes the predicted CTR, and  $y_{\text{XGBoost}}$ ,  $y_{\text{FM}}$ , and  $y_{\text{DNN}}$  are the outputs of the XGBoost, FM, and deep components, respectively.

**2.1. XGBoost Component.** The XGBoost component is a scalable machine learning system for tree boosting [2]. In XGBoost, feature selection and combination are automatically performed to generate new discrete feature vectors as the input of the LR model. The depth of a decision tree determines the dimension of the feature intersection. For example, if the depth of the decision tree is four, then the final number

of the leaf node is the number of orders (three order) of feature interactions. We use XGBoost to capture three-order feature interaction and perform feature selection among features. The objective function of the XGBoost is as follows:

$$\text{Obj} = \sum_{i=1}^N l(\hat{y}_i, y_i) + \sum_{k=1}^K \Omega(f_k). \quad (2)$$

XGBoost uses the following forward distribution algorithm:

$$\hat{y}_i^{(t)} = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i), \quad (3)$$

$$\text{Obj}^{(t)} = \sum_{i=1}^N l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) + C,$$

where  $\hat{y}_i^{(t)}$  is the predicted value of the time  $t$  of the iteration, that is, the predicted result of sample  $x_i$  by  $t$  trees and  $\hat{y}_i^{(t-1)}$  is the predicted value of the current  $(t-1)$ th iteration. Thus, when the model is initialized, the model has no tree, and the predicted result is a constant. Each iteration adds a new tree to the model, and the loss function then changes correspondingly. In addition, the training of  $(t-1)$ th trees is completed when the  $(t)$ th tree is added.

**2.2. FM Component.** The FM component is used to learn feature interactions [1]. FM models can capture two-order feature interactions as the inner product of respective feature latent vectors.

$$\hat{y}(x) = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle v_i, v_j \rangle x_i x_j, \quad (4)$$

where  $v_i, v_j$  denotes the latent vector. Each cross-term parameter  $w_{ij}$  is expressed by the inner product  $\langle v_i, v_j \rangle$  of the latent vector. The objective function of FM is as follows:

$$\text{Obj}^{(t)} = \sum_{i=1}^N l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) + C. \quad (5)$$

**2.3. Deep Component.** The deep component is used to learn high-order (more than three orders) feature interactions. The original features are initially embedded such that the features of different fields are mapped to the same dimension of the embedding space. Similarly, the dimension of the implicit vectors is  $k$ . Here, we set two layers for the deep component, and the entire DNN component is then computed as follows:

$$\begin{aligned} a^{(0)} &= [e_1, e_2, \dots, e_m], \\ a^{(1)} &= \sigma(W^{(0)} a^{(0)} + b^{(0)}), \\ a^{(2)} &= \sigma(W^{(1)} a^{(1)} + b^{(1)}), \\ \hat{y}_{\text{DNN}} &= W^{(3)} a^{(2)} + b^{(3)}, \end{aligned} \quad (6)$$

where  $e_m$  denotes the embedding of discrete features and  $\hat{y}_{\text{DNN}}$  is the prediction of DNN for the CTR of mobile

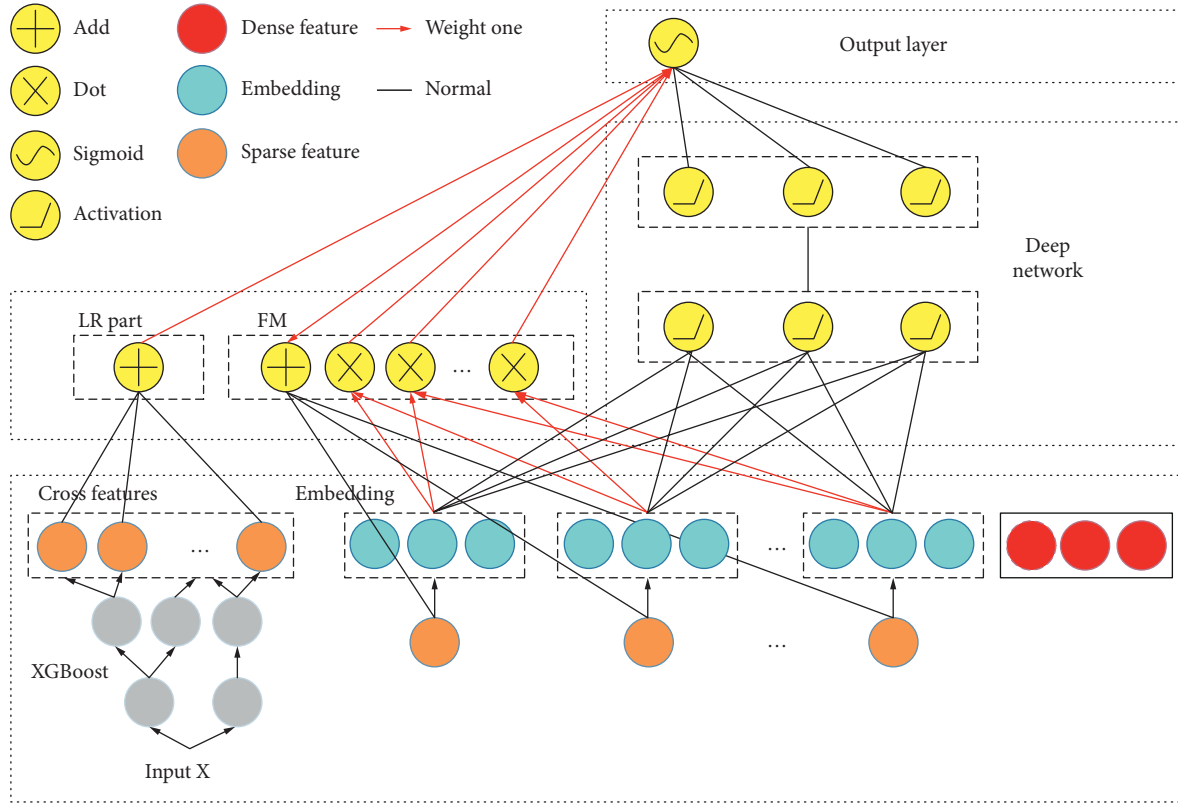


FIGURE 1: XGBDeepFM model.

advertising. After the feature selection of the XGBoost, the FM and deep components share the same feature embedding.

### 3. Experiment

**3.1. Datasets.** The dataset used in this study contains the O2O mobile ad data from a mobile Internet platform, which provides users with local life service information. The dataset covers offline scenes, such as catering, supermarkets, convenience stores, takeout, beauty salons, and cinemas. The platform not only provides rich ad information but also offers users' explicit and implicit behavior information. Such abundance of data brings great convenience in CTR prediction. The original experimental dataset contains attributes such as shop information, users' payment log, and users' browsing log in 2016 (see Tables 1 and 2).

Mobile contextual ad CTR has a significant relationship with weather. Thus, we also crawl weather information from a weather platform named <http://WunderGround.com>. The platform is a reliable source of historical weather forecast information on a global scale. In this study, 4,369,918 precise historical weather data of 122 cities on day and hour levels are crawled, as shown in Tables 3 and 4, respectively.

**3.2. Evaluation Metrics.** We use the area under the ROC curve (*AUC*) as our evaluation metric because it is not bias on the size of test or evaluation data. *AUC* measures the likelihood that given two random points, one from the

positive and one from the negative class, the classifier will rank the point from the positive class higher than the one from the negative one. The larger the *AUC* is, the more accurate the CTR prediction of mobile advertising will be.

$$AUC = \sum_{i \in (P+N)} \frac{(TPR_i + TPR_{i-1})(FPR_i - FPR_{i-1})}{2} \quad (7)$$

**3.3. Feature Selection by XGBoost.** A benefit of using XGBoost is that, after the boosted trees are constructed, importance scores that indicate how useful or valuable each feature is in the construction of the boosted decision trees within the model can be easily obtained. Thus, we choose XGBoost because this model is easily interpretable by human experts. Moreover, the depth of the decision tree can decide the order of feature interaction, which can make up for the FM and DNN components. We plot the feature importance calculated by the XGBoost model, as shown in Figure 2.

The ranking results of the feature importance show that contextual features are of high importance. For example, from the perspective of temporal features, the week ranks fourth in the importance of the model; from the perspective of geographical features, geographical location features rank ten; and from the perspective of temperature contextual features, pressure and body temperature, which are important features, rank second. This feature selection focuses on the integration of mobile ad bilateral factors (i.e., ad and user factors) and contextual factors, as shown in Table 5.

TABLE 1: Mobile ad shop information.

Field	Sample	Description
shop_id	000001	Shop id
city_name	Shenzhen	City name
location_id	001	Neighbour shops have the same location id
per_pay	3	Average pay
Score	1	Shop score
comment_cnt	2	Users' comment number
shop_level	1	Shop level
cate_1_name	Food	Category name
cate_2_name	Snack	Category name
cate_3_name	Other	Category name

TABLE 2: User payment behavior/browsing behavior.

Field	Sample	Description
user_id	0000000001	User id
shop_id	000001	Shop id
time_stamp	2016-10-10 11:00:00	Payment/browsing timestamp

TABLE 3: Weather data (day level).

Field	Sample	Description
city_name	Shenzhen	City name
Date	2016-01-01	Date
d_weather	Cloud	Weather
d_temp	11°C/4°C	Highest/lowest temperature
d_wind	SE	Wind direction

TABLE 4: Weather data (half-hour level).

Field	Sample	Description
city_name	Shenzhen	City name
Date	2016-01-01	Date
Time	12:30 AM	Time
h_temp	26.0	Temperature
h_bodytemp	24.0	Body temp
h_dew	24.0	Dew
h_humidity	89%	Humidity
h_pressure	1011	Pressure
h_visibility	10.0	Visibility
h_wind_dir	South	Wind direction
h_wind_speed	18.5	Wind speed
h_gust_speed	19	Gust speed
h_condition	Clouds	Weather condition

**3.4. Model Comparison.** We initially compare the performance of each component of the model (i.e., first-order linear, second-order FM, DNN, and XGB components) and its combination under the optimal settings. Then, we compare our proposed method with other models, namely, W&D, FNN, PNN, and XDeepFM.

Figure 3 presents the *AUC* results of in-model comparisons. We compare the predictive performance of different models and observe the following. First, the models with the FM component are better than those without it. The linear model, linear + DNN model, linear + XGB model, and linear + DNN + XGB model are improved by 0.1121, 0.0018,

0.0587, and 0.0003, respectively, after adding the FM component. Second, the models with the DNN component are better than those without it. The linear model, linear + FM model, linear + XGB model, and linear + FM + XGB model are improved by 0.2199, 0.1086, 0.1157, and 0.0573, respectively, after adding the DNN component. Third, the models with the XGB component are better than those without it. The linear model, linear + DNN model, linear + FM model, and linear + FM + DNN improved by 0.1049, 0.0007, 0.0515 and 0.0003, respectively, after adding the XGB component. The experimental results show that the FM part, DNN part, and XGB part have significant gains on the

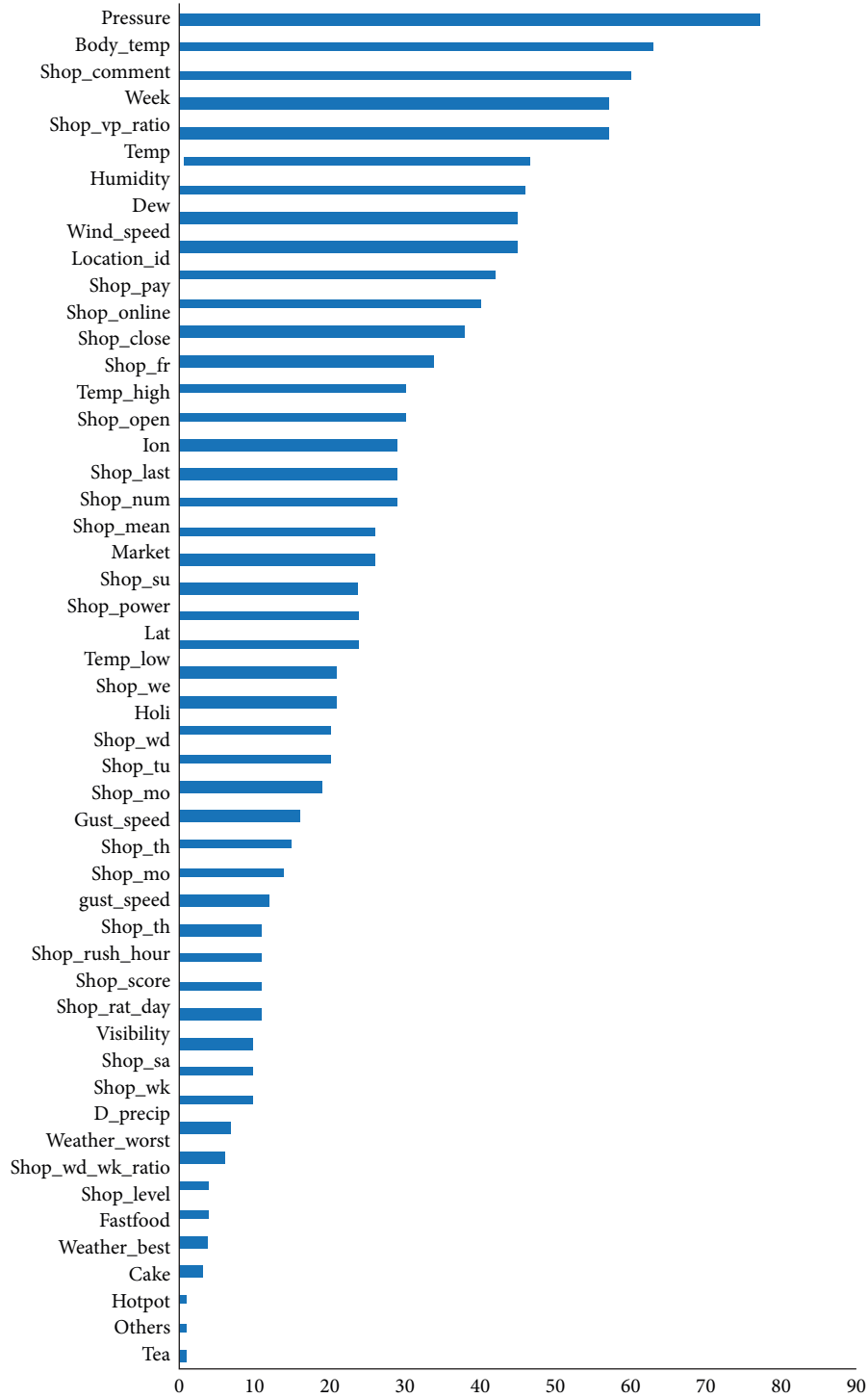


FIGURE 2: Feature importance ranking.

TABLE 5: Feature selection.

Category	Dimension	Description
Bilateral	Ad	Including price level, sales volume, ratings, categories, click-through rate
	User	Including user's category preference, price preference, location preference
	Time	Including seasonal characteristics, weekends, holidays, hour peak
Contextual	Geography	Including the user's active geographic location, location
	Weather	Including weather conditions, temperature, humidity, wind speed

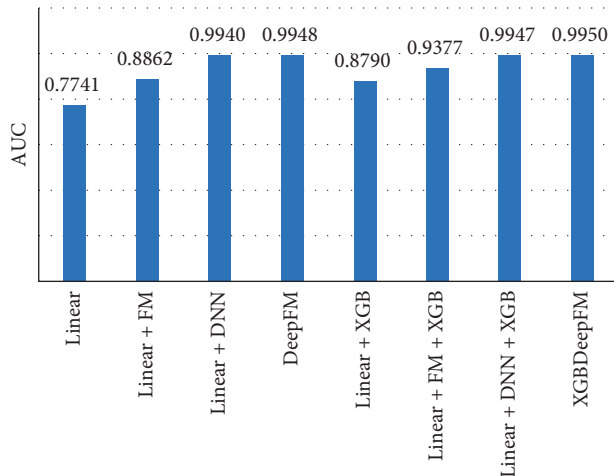


FIGURE 3: Prediction performance comparison of subcomponents.

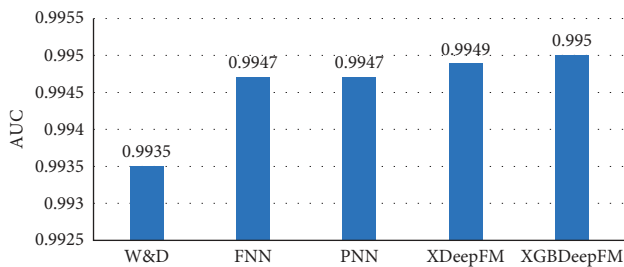


FIGURE 4: Prediction performance comparison.

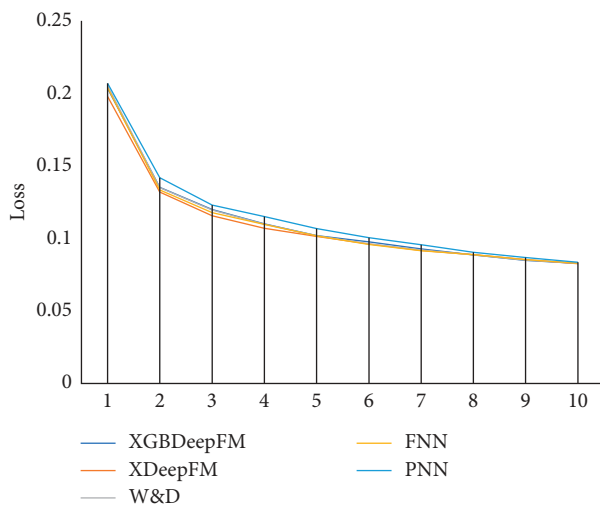


FIGURE 5: Convergence time comparison of XGBDeepFM with the deep learning model.

model, which are expected. The XGBDeepFM model has a strong information capacity for the CTR prediction of mobile advertising by integrating bilateral and contextual factors. Therefore, the prediction performance of the eight models is the best.

From the experimental results shown in Figure 4, XGBDeepFM is superior to all depth CTR models in terms

of the AUC index. XGBDeepFM is 0.0015, 0.0003, 0.0003, and 0.0001 higher than W&D, FNN, PNN, and XDeepFM, respectively. Overall, the following results are obtained (Figures 3 and 4):

- (1) Learning feature interactions instead of learning only linear features improves the performance of CTR prediction
- (2) Learning low- and high-order feature interactions simultaneously contributes to CTR prediction
- (3) Learning more important features based on the XGBoost model can improve the performance of a CTR prediction model

Figure 5 shows the comparison results of the convergence time of different models. The results show that the convergence speed of the XGBDeepFM algorithm is faster than that of PNN and FNN, only next to XDeepFM; the loss is the lowest after the 10th round of training.

## 4. Conclusions

In CTR prediction, the contextual features and interactions among ad, user, and contextual features are key factors that can affect the prediction performance. In this study, we propose the XGBDeepFM model. We initially include information on contextual features to improve the prediction accuracy from the perspective of time, geography, and weather. Then, a feature selection process is conducted to obtain important features. Low- and high-order features are obtained using the proposed XGBDeepFM model. We conduct extensive experiments to compare the effectiveness and efficiency of XGBDeepFM with other methods. Our experiment results demonstrate that (1) XGBDeepFM outperforms the state-of-art models in terms of AUC, and (2) the efficiency of XGBDeepFM outperforms most deep neural network models.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

This work was supported by the Natural Science Foundation of China (grant nos. 71831005 and 71472056) and Shenzhen Key Research Base of Humanities and Social Sciences.

## References

- [1] H. Guo, R. Tang, Y. Ye et al., "DeepFM: a factorization-machine based neural network for CTR prediction," 2017, <http://arxiv.org/abs/1703.04247>.
- [2] H. T. Cheng, L. Koc, J. Harmsen et al., "Wide & deep learning for recommender systems," in *Proceedings of the 1st Workshop*

- on *Deep Learning for Recommender Systems*, pp. 7–10, Boston MA USA, September 2016.
- [3] F. García-Sánchez, R. Colomo-Palacios, and R. Valencia-García, “A social-semantic recommender system for advertisements,” *Information Processing & Management*, vol. 57, no. 2, p. 102153, 2020.
  - [4] Q. Wang, F. Liu, S. Xing, and X. Zhao, “Research on CTR prediction based on stacked autoencoder,” *Applied Intelligence*, vol. 49, no. 8, pp. 2970–2981, 2019.
  - [5] J. Chen, B. Sun, H. Li et al., “Deep Ctr prediction in display advertising,” in *Proceedings of the 24th ACM International Conference on Multimedia*, pp. 811–820, Amsterdam, The Netherlands, October 2016.
  - [6] M. Karimi, D. Jannach, and M. Jugovac, “News recommender systems-survey and roads ahead,” *Information Processing & Management*, vol. 54, no. 6, pp. 1203–1227, 2018.
  - [7] P. Resnick, N. Iacovou, M. Suchak et al., “GroupLens: an open architecture for collaborative filtering of netnews,” in *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, pp. 175–186, Chapel Hill, NC, USA, October 1994.
  - [8] W. Carrer-Neto, M. L. Hernández-Alcaraz, R. Valencia-García, and F. García-Sánchez, “Social knowledge-based recommender system. Application to the movies domain,” *Expert Systems with Applications*, vol. 39, no. 12, pp. 10990–11000, 2012.
  - [9] F. M. Belém, A. G. Heringer, J. M. Almeida, and M. A. Gonçalves, “Exploiting syntactic and neighbourhood attributes to address cold start in tag recommendation,” *Information Processing & Management*, vol. 56, no. 3, pp. 771–790, 2019.
  - [10] L. Tang, Y. Jiang, L. Li et al., “Ensemble contextual bandits for personalized recommendation,” in *Proceedings of the 8th ACM Conference on Recommender Systems*, pp. 73–80, Foster City, SV, USA, October 2014.
  - [11] C. C. Aggarwal, *An Introduction to Recommender systems*, pp. 1–28, Springer, Cham, Switzerland, 2016.
  - [12] T. Rangunathan, S. K. Battula, V. Jorika, C. Mounika, A. U. Sruthi, and M. D. Vani, “Advertisement posting based on consumer behaviour,” *Procedia Computer Science*, vol. 50, pp. 329–334, 2015.
  - [13] H. B. McMahan, G. Holt, D. Sculley et al., “Ad click prediction: a view from the trenches,” in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1222–1230, Chicago, IL, USA, August 2013.
  - [14] S. Rendle and L. Schmidt-Thieme, “Pairwise interaction tensor factorization for personalized tag recommendation,” in *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, pp. 81–90, New York NY, USA, February 2010.
  - [15] Y. Zhang, H. Dai, C. Xu et al., “Sequential Click Prediction for Sponsored Search with Recurrent Neural networks,” in *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, Québec City, Canada, July 2014.
  - [16] Y. Qu, H. Cai, K. Ren et al., “Product-based neural networks for user response prediction,” in *Proceedings of the 2016 IEEE 16th International Conference on Data Mining*, IEEE, Barcelona, Spain, pp. 1149–1154, December 2016.