

Research Article

Advertisement Click-Through Rate Prediction Based on the Weighted-ELM and Adaboost Algorithm

Sen Zhang,^{1,2} Qiang Fu,^{1,2} and Wendong Xiao^{1,2}

¹School of Automation & Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China

²Key Laboratory of Knowledge Automation for Industrial Processes, Ministry of Education, Beijing 100083, China

Correspondence should be addressed to Sen Zhang; zhangsen@ustb.edu.cn

Received 13 July 2017; Accepted 4 October 2017; Published 9 November 2017

Academic Editor: Wenbing Zhao

Copyright © 2017 Sen Zhang et al. 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.

Accurate click-through rate (CTR) prediction can not only improve the advertisement company's reputation and revenue, but also help the advertisers to optimize the advertising performance. There are two main unsolved problems of the CTR prediction: low prediction accuracy due to the imbalanced distribution of the advertising data and the lack of the real-time advertisement bidding implementation. In this paper, we will develop a novel online CTR prediction approach by incorporating the real-time bidding (RTB) advertising by the following strategies: user profile system is constructed from the historical data of the RTB advertising to describe the user features, the historical CTR features, the ID features, and the other numerical features. A novel CTR prediction approach is presented to address the imbalanced learning sample distribution by integrating the Weighted-ELM (WELM) and the Adaboost algorithm. Compared to the commonly used algorithms, the proposed approach can improve the CTR significantly.

1. Introduction

With the development of the network technology and the communication technology, the Internet and the mobile Internet have been developed rapidly. Due to the popularity of smart phones, a variety of the mobile phone applications are invented. It is a niche market where the advertisers and the advertising companies pay more attention to the click-through rate (CTR) in the online advertising products. Usually the online advertising can be done in two different ways: one is the website search based advertising, which specifically refers to the searching engine depending on the user's key words that target the advertising content and the advertising spot. The other one is the real-time bidding (RTB) advertising, in which the advertising supplier platform provides no longer the advertising spot, but the specific users who visited the advertisement spot. The RTB advertisements enlarge the online advertising's directivity and accuracy [1].

Currently, there exists many research works on CTR prediction for Internet advertising. Menon et al. [2] proposed the maximum likelihood algorithm to estimate the parameters of the CTR probabilistic model. But this model can only be applied to the existing advertisements rather than

the new advertisements. Richardson et al. [3] proposed the logic regression model to learn the CTR prediction model for searching advertising with the model features including the number of the keywords, the position of the figures in the page, and the other characteristics of the advertisements. Chapelle [4] proposed a stochastic regression approach based on the rate estimation machine learning framework for the Yahoo! to solve the CTR prediction problem by using four features as the model inputs. The norm-2 regularization term is added in the logistic regression model. This method can produce a sparser model to increase the number of the nonzero parameters to avoid the overfitting problem. Shao [5] proposed a high-level feature representation and a click-by-point prediction method based on the deep network that combines the high-level features and the basic features by using deep neural network model.

Most existing work on CTR prediction is focused on searching advertising that is seriously dependent on the keyword and the user input. With the development of the intelligent terminals and the mobile Internet, RTB advertising is increasing rapidly. More and more advertisers are in favor of the RTB advertising which will become the main trend of the Internet advertising in the future. At the same time, the

TABLE 1: Description of the experimental dataset.

Attribute name	Data type	Attribute explanation
push_time	Timestamp	Time of bidding request
u_id	String	User ID
exchange_id	Int	Advertising Exchange Platform ID
c_id	String	Advertising Creative ID
space_id	String	Advertising Position ID
area_id	Int	Area ID
media_id	Int	Media ID
advertiser_id	Int	Advertiser ID
policy_id	Int	Policy ID
user_agent	String	Agent of browser
user_ip	String	User ip
if_click	Int	If click
if_show	Int	If show
price_base	Double	Lowest price for bidding the advertising position
price_win	Double	Price for winning the advertising position
url	String	URL

research work on the RTB CTR prediction is still at the beginning stage.

In this paper, we will study the novel big data based online CTR prediction problem by incorporating RTB advertising with user profile system. A novel CTR prediction approach will be presented by integrating the Weighted-ELM (WELM) and the Adaboost algorithm to address the imbalanced learning sample distribution. We will perform the experiments using real advertising datasets to verify the effectiveness of the proposed approach.

2. The Experimental Dataset and the Evaluation Criteria

In this section, the experimental dataset and the evaluation criteria used in this study named Area Under Curve (AUC) will be briefly described.

The experimental dataset used in this paper for CTR prediction is the original data log provided by a domestic advertising company in China. There are 16 attributes in the original data log, with the details shown in Table 1.

2.1. User Profile. Since the advertising log has large amount of data, we divide the above 16 attributes into 4 categories: the user's characteristics, the temporal characteristics, the ID characteristics, and the numerical characteristics.

2.1.1. The User's Characteristics. In early practice, when the demand side platform receives the bidding request from the advertising agent, normally the user's information is not analyzed and all users were used for advertising. It is proved that this way of the information delivery cannot achieve the desired results as the u_id and media_id attributes used in the

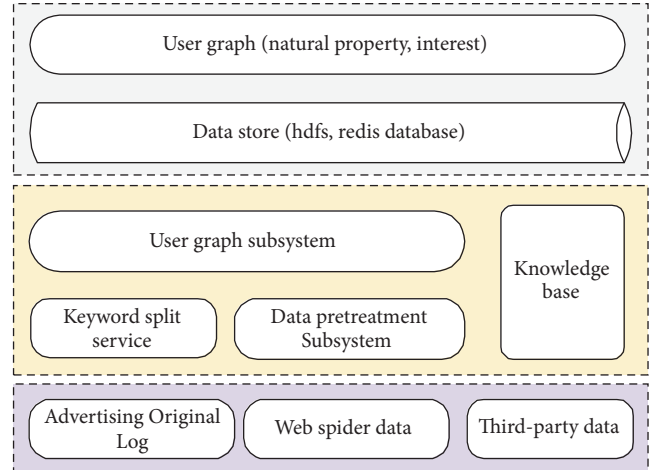


FIGURE 1: The block diagram of the user profile system.

approach cannot cater to the users' interest. Thus the primary task is to establish the user profile system to obtain the user's age, gender, and interest preference for CTR prediction. The overall structure of the system is shown in Figure 1.

The user profile system mainly includes the following functions:

- (i) Data pretreatment subsystem: take the responsibility of cleaning and preprocessing the advertising log data;
- (ii) Keyword split service: take the responsibility of segmenting the irregular text;
- (iii) Knowledge base: take the responsibility of providing the related mapping tables;
- (iv) User graph subsystem: the most important part of the user graph system: take the responsibility of integrating various parts of the data to build a user graph;
- (v) Data storage subsystem: take the responsibility of storing the results of the user graph.

The output of the user graph system includes the user's age, gender, and interest preference. The users' characteristics are obtained by using i_id attribute to match the output of the user graph system.

2.1.2. The Time Characteristics. The time characteristics include the field of push_time in the log which represents the time of the ads request. According to the historical data, the users have different interests at different time periods, so the probability of a click behavior is also different. Based on this judgment, we split one day into six time periods which are late-night, morning, lunch time, afternoon, dinner time, and evening. The entire time information is organized by a six-dimensional vector. The six periods of time are shown in Table 2.

2.1.3. The ID Characteristics. The ID characteristics in the dataset include the u_id, the advertiser_id, the media_id, the

TABLE 2: Information of a whole day.

Period name	Period
T1 (late-night)	00:00~06:00
T2 (morning)	06:00~11:00
T3 (lunchtime)	11:00~13:00
T4 (afternoon)	13:00~18:00
T5 (dinnertime)	18:00~20:00
T6 (evening)	20:00~23:00

area_id, the c_id, the policy_id, and the exchange_id. There are a lot of ID attributes in the RTB advertising logs. If we do not have the filtering process of the characteristics, we would obtain a vector whose dimension may be up to several hundred thousands which increases the computational complexity seriously. Therefore, it is necessary to reduce the dimensionality of the feature space. We apply the method in [3] to remove the needless ID attributes that have no impact or little impact on the click-through rate.

2.1.4. The Numerical Characteristics. Attributes in the dataset, such as the price_base, the price_win, the URL, and the u_ip, affect the advertising's CTR as well. Take the price_win for example, if the value is 0, it indicates that the advertising is not a successful bidding. If the value is nonzero, the different values reflect that the value of the advertising clicking is different. It is usually considered that the larger the value is, the better the advertising position is and the greater the probability of the clicking is. Therefore the numerical attributes need to be added to the feature vector.

In this paper, we adopted the maximum and minimum normalization method to normalize each characteristic to the value between 0 and 1.

2.2. Area Under Curve (AUC). The prediction of the CTR is a binary classification problem while the proportion of the positive and negative samples is extremely uneven. In the actual advertising, the proportion of the positive and the negative samples is about 3 : 1000 or even lower. The samples are distributed in different categories unevenly, so the evaluation index of accuracy is not a good criterion to judge the performance of the classifier.

In this paper, AUC is adopted to measure the effect of the CTR prediction. In the process of calculating the AUC, the related curve is called ROC curve (receiver operating characteristics) [6]. Traditional ROC curve is used in medical field. Currently it is often used in the field of data mining, machine learning, and pattern recognition.

When the ROC curve is drawn, the horizontal coordinate is FPR (False Positive Rate) and the vertical coordinate is TPR (True Positive Rate). The values of FPR and TPR can be calculated according to the formula (1).

$$\begin{aligned} \text{FPR} &= \frac{\text{FP}}{\text{N}} = \frac{\text{FP}}{\text{FP} + \text{TN}} \\ \text{TPR} &= \frac{\text{TP}}{\text{P}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \end{aligned} \quad (1)$$

TABLE 3: The AUC value of each ID attribute.

ID attribute	exchange_id	area_id	media_id	advertiser_id
AUC value	0.5137	0.5412	0.5930	0.5271

In (1), TP represents the fact that the samples are positive and the algorithm recognizes them as the positive samples; FP represents the fact that the samples are negative and the algorithm recognizes them as the positive samples; FN represents the fact that the samples are positive and the algorithm recognizes them as the negative samples; TN represents the fact that the samples are negative and the algorithm recognizes them as the negative samples [7].

It is obvious that if there are more users to click an advertisement, the rank of this advertisement will be in the front and the area under the ROC curve is larger which indicates that the performance of the advertising is better.

As an example, we draw the receiver operating characteristics (ROC) curves for the exchange_id, the area_id, the media_id, and the advertiser_id by the Weighted-ELM. Each AUC value of the curve is shown in Table 3.

From Table 3, we can see that the AUC values of the exchange_id and the advertiser_id are almost 0.5, which have no difference from the random results. This phenomenon has something related to the characteristics of the RTB advertising. The RTB advertisers do not want their own click conversion data to be used to optimize the other advertisers' effectiveness.

Compared to the AUC value of the advertiser_id, the AUC value of the media_id is increased slightly and up to 0.60. This case is related to the user's interest and the media_id can reflect the user's interest. If the users visit a few apps frequently, the probability of clicking the ads would be increased.

3. The CTR Assessment

In this section, the ELM algorithm will be discussed, which will be used in the prediction of the CTR. Compared with the traditional classification algorithms SVM and BP, the ELM has the advantage of fast learning speed and accurate estimation results with easily setting the weights. Based on these advantages, the ELM algorithm has been developed rapidly since it was proposed several years ago. Because the proportion of the positive and the negative samples is extremely uneven, we proposed the Weighted-ELM algorithm to solve the problem in the next subsection. Because the ELM is the basis of the Weighted-ELM algorithm, we will firstly describe the original ELM in the following.

3.1. The ELM Algorithm. In recent years, Huang et al. [8–10] and the other scholars proposed a fast algorithm of single-hidden layer feedforward neural network named extreme learning machine (ELM) [11, 12]. The specific structure of the ELM algorithm is shown in Figure 2.

The input weights and the bias of the hidden node in the ELM are chosen randomly. They do not need a series of iterative algorithm, which greatly saves the training time of

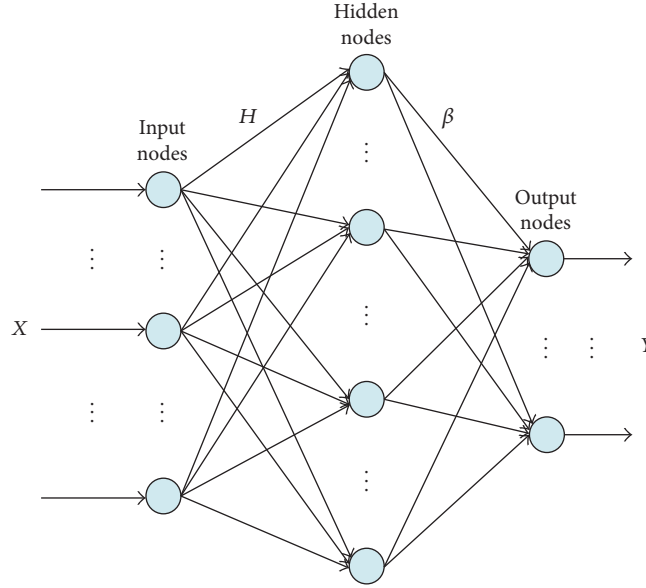


FIGURE 2: The structure of the ELM neural network.

the neural network. The output weights of the ELM are obtained by minimizing the squared error loss function to get the least square solution. Thus the process of determining the neural network parameters is very simple that saves much time of adjusting the parameters.

The basic idea of the ELM algorithm is as follows.

The training sample set is given as $\{(X, T) \mid X = [X_1, X_2, \dots, X_N]^T, T = [t_1, t_2, \dots, t_N]\}$, where the matrix X is the input matrix of the neural network and the matrix T is the actual output value of the training sample set. From the neural network with L hidden nodes, we can get

$$\sum_{i=1}^L \beta_i \cdot G(a_i \cdot X_j + b_i) = O_j, \quad j = 1, 2, \dots, N. \quad (2)$$

In this equation, $G(ax + b)$ is the neural network hidden layer node activation function. Usually it is sig, sin, hardlim, or tribas function; a_i is the connection weights between the i th hidden layer node and the input nodes; b_i is the bias of the i th hidden node; β_i is the connection weights between the i th hidden layer nodes and the output node.

In the practical application of the algorithm, the output value of the network is equal or near to the actual output value. If the sample set and the neural network structure are close to the target value T with the zero error, we can get $\|H\beta - T\| = 0$. The formula of the ELM algorithm can be abbreviated as

$$H\beta = T, \quad (3)$$

where H is the output matrix of the neural network hidden nodes and β is the output weight matrix between the hidden layer nodes and the output layer node.

The main idea of the algorithm is how to get the output weight matrix β to make the training error $\|H\beta - T\|^2$ and

the output weight matrix $\|\beta\|$ minimum. That means how to make the following equation's value minimum:

$$J = (H\beta - T)^T (H\beta - T) \quad (4)$$

$$\beta = H^T T,$$

where H^Γ is the generalized inverse matrix of H . If $H^T H$ is nonsingular, $H^\Gamma = (H^T H)^{-1} H^T$. If HH^T is nonsingular, $H^\Gamma = H^T (HH^T)^{-1}$. If H is not full column rank, β could be obtained by the singular value decomposition (SVD) [5, 13].

3.2. The Weighted-ELM Algorithm. The basic ELM algorithm is very useful for many problems. However, there exist a lot of classification problems whose samples are imbalance, such as the advertising click rate problem. In order to solve the problem of the sample imbalance in the classification, Xu et al. proposed the Weighted-ELM algorithm [14].

The objective function of the ELM algorithm is

$$L_{\min} = \frac{1}{2} \|\beta\|^2 + \frac{C}{2} \sum_{i=1}^N \xi_i^2. \quad (5)$$

In this equation, the condition is satisfied: $h(x_i)\beta = t_i - \xi_i, i = 1, 2, \dots, N$. The first half of formula (5) is called the structural risk, and the latter part is called the empirical risk.

The objective function of the Weighted-ELM algorithm is

$$L_{\min} = \frac{1}{2} \|\beta\|^2 + \frac{C}{2} \mathbf{W} \sum_{i=1}^N \xi_i^2, \quad (6)$$

where \mathbf{W} is an $N \times N$ diagonal matrix and the value of the matrix \mathbf{W} is related to each training sample. Generally, if \mathbf{x}_i belongs to a few classes, the corresponding \mathbf{W}_{ii} should be

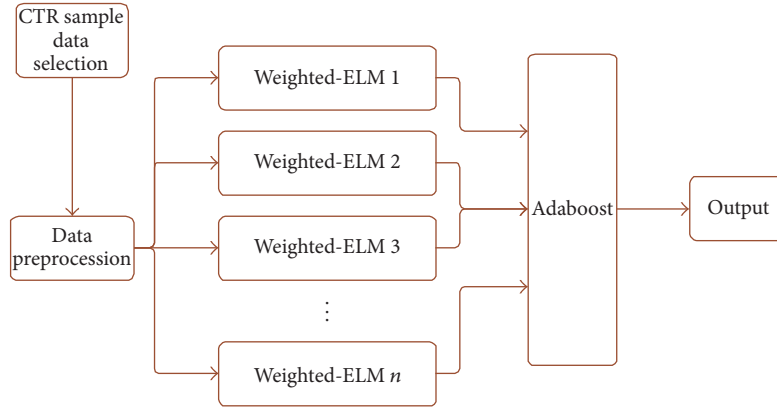


FIGURE 3: The flowchart of the WELM-Adaboost algorithm.

given a relatively large weight. There are two methods for the value of w . The first method is shown in

$$W_{ii} = \frac{1}{\#(t_i)} \quad (7)$$

$$W_{ii} = \begin{cases} \frac{k}{\#(t_i)} & \text{if } t_i > \text{avg}(t_i) \\ \frac{1}{\#(t_i)} & \text{if } t_i \leq \text{avg}(t_i); \end{cases} \quad (8)$$

the second method is as follows.

The process of training ELM is equivalent to solving the following problem:

$$L_{\text{ELM}} = \frac{1}{2} \|\beta\|^2 + \frac{1}{2} \text{CW} \sum_{i=1}^N \xi_i^2 - \sum_{i=1}^N \alpha_i (\mathbf{h}(\mathbf{x}_i) \beta_i - t_i + \xi_i). \quad (9)$$

Similar to the original ELM, β is also solved in two ways:

When N is small,

$$\beta = \mathbf{H}^T \left(\frac{\mathbf{I}}{C} + \mathbf{W}\mathbf{H}\mathbf{H}^T \right)^{-1} \mathbf{W}\mathbf{T}. \quad (10)$$

When N is large,

$$\beta = \left(\frac{\mathbf{I}}{C} + \mathbf{H}^T \mathbf{W}\mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{W}\mathbf{T}. \quad (11)$$

The output of the Weighted-ELM classifier can be given by

$$f(x) = \mathbf{h}(x) \beta. \quad (12)$$

4. WELM-Adaboost Algorithm

This paper constructs the advertisement click rate prediction model by the proposed WELM-Adaboost algorithm which can adjust the weight of the data distribution.

4.1. Adaboost Algorithm. Adaboost algorithm is one of the typical applications of the Boosting algorithm. The Adaboost algorithm chooses the very important features to construct a series of weak classifiers and cascade these weak classifiers to form a stronger classifier. The advantage of this algorithm is that it uses the weighted training data instead of the randomly selected training samples. It combines the weak classifiers and uses the weighted voting mechanisms instead of the average voting mechanism.

4.2. The Advertisement Click Rate Prediction Model Based on the WELM-Adaboost. In this paper, the Weighted-ELM is used as a weak predictor, and the weight distribution of each sample is adjusted by using the Adaboost algorithm to obtain multiple Weighted-ELM classifiers. These classifiers are combined into a strong classifier [14].

The advertisement click rate prediction process based on the WELM-Adaboost algorithm is shown in Figure 3.

The detailed steps of the algorithm are as follows:

- (1) From the sample data, randomly select N sample data as the training data. According to the positive and the negative samples of the distribution ratio, initialize the weights of each training sample.
- (2) For each iteration $m = 1 : M$, where M is the total number of the weak classifiers, the algorithm will repeat the following steps from (a) to (e):

- (a) Apply the training samples to a classifier $\text{ELM}_m(x)$ with the initial sample weight w_i ;
- (b) Calculate the weight prediction error from the weights of the $\text{ELM}_m(x)$ whose results are misclassified samples; the weight prediction error is calculated according to

$$\text{err}_m = \frac{\sum_{i=1}^N w_i I(y_i \neq \text{ELM}_m(x_i))}{\sum_{i=1}^N w_i}. \quad (13)$$

- (c) Calculate the weight of the sequence α_m of the $\text{ELM}_m(x)$ according to its classification performance:

TABLE 4: The statistics of the experimental dataset.

Dataset	Impression_num	Click_num	Click/impression
Training set	180823	544	0.0030
Test set	77496	252	0.0032
total	258319	796	0.0031

$$\alpha_m = \frac{1}{2} \ln \left(\frac{1 - \text{err}_m}{\text{err}_m} \right). \quad (14)$$

(d) The weight of the new training sample is adjusted according to the calculated sequence weight α_m :

$$w_i = w_i \cdot \exp(\alpha_m \cdot I(y_i \neq \text{ELM}_m(x_i))). \quad (15)$$

(e) Renormalize the sample weight.

(3) After M iterations, the M -group weak predictors are obtained. These weak predictors are merged into the final strong predictor $C(x)$:

$$C(x) = \arg \max_k \sum_{m=1}^M \alpha_m \cdot I(\text{ELM}_m(x) = k), \quad (16)$$

where k is the number of the categories of the samples.

5. The Experimental Results

The experimental dataset used in this paper is the RTB advertisement raw log data provided by a domestic advertisement company in Beijing, China. Since the data is too large and the positive (or the negative) samples are seriously imbalanced, we randomly extract 1‰ of the data as the experimental data from the log. Click samples are recorded as positive; the other (nonclick) samples are negative. The proportion of the positive and the negative samples of the experimental data is almost 3:1000 which is a typical unbalanced data set. The statistics of the experimental data is shown in Table 4. In the table, Impression_n means the number of the nonclick samples and the Click_num means the number of the click samples.

5.1. The CTR Prediction Model. From the above feature extraction process, we can conclude that the CTR of the RTB advertisement has a great relationship with the users' interest and the basic attributes. It has a little relationship with most of the ID characteristics. Finally, we select the temporal characteristics and the user characteristics like *media_id*, *area_id*, *price_base*, and *price_win* as the input of the prediction model based on the proposed method.

It is necessary to explore the influence of the number of the hidden nodes and the activation function on the speed and the accuracy of the ELM algorithm.

The ELM algorithm provides four kinds of activation functions. From Figure 4, we can know that when the number

TABLE 5: Overall CTR estimation AUC performance.

	LR	SVM	ELM
1:5	0.852	0.846	0.951
1:10	0.803	0.818	0.838
1:20	0.751	0.761	0.840
1:50	0.732	0.712	0.839
1:100	0.625	0.639	0.682
1:150	0.507	0.510	0.508

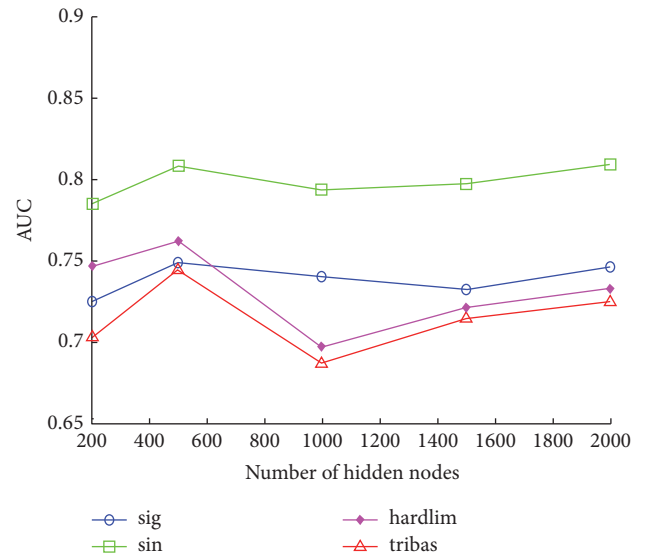


FIGURE 4: The trend of AUC value with different activation function.

of the hidden nodes is the same and the activation function is sine function, the AUC value is higher than the other three types of the activation functions about 5%. In addition, the training speed of the sine function is slower than the sigmoid function and the tribas function, but faster than the hardlim function. Considering the training time and the equipment cost, the number of the hidden nodes is set to 500, and the activation function is set to sine function.

5.2. The Comparison of the Algorithms' Performance. We select logistic regression (LR) model and support vector machine (SVM) model as the comparison methods which are commonly used in other papers, and AUC values of three algorithms are shown in Table 5.

Table 5 shows that the performance of ELM is better than LR and SVM on all the tested datasets, which shows that we have chosen the reasonable characteristics and ELM algorithm is effective as well.

Finally, we selected the traditional ELM algorithm and the Weighted-ELM algorithm as a contrast method when the positive and the negative samples' ratios are set with different proportions; the trend of the AUC results of the three algorithms is shown in Figure 5.

It can be seen from Figure 5 that when the positive and the negative sample ratio is 1:5, the three algorithms' AUC values can reach 0.9 or more. When the positive and the negative

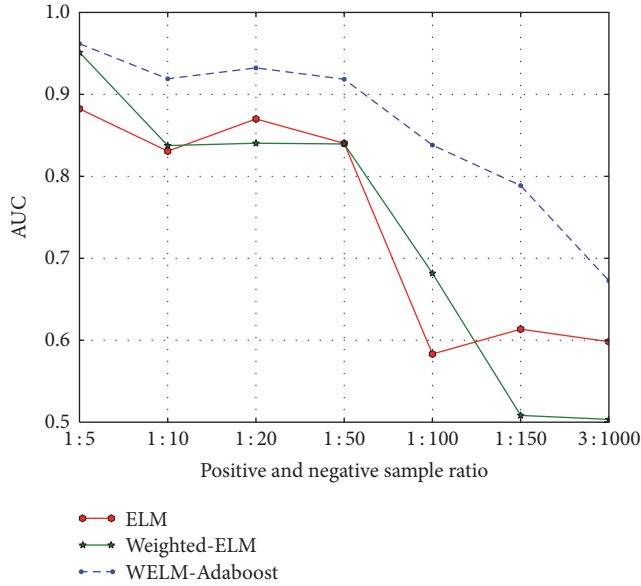


FIGURE 5: The comparison results of AUC of the three algorithms.

sample ratio is 1:50, the AUC value of the WELM-Adaboost algorithm is still above 0.9, but the AUC values of the ELM algorithm and the Weighted-ELM algorithm reduced to 0.84. With the increasing ratio of the sample proportion of the positive and the negative samples, the AUC values of the three algorithms show a decreasing trend, but the AUC value of the WELM-Adaboost algorithm is obviously higher than that of the other two algorithms. The proposed WELM-Adaboost algorithm has a better performance than the other two methods.

The results are shown in Table 6.

For the WELM-Adaboost algorithm, this algorithm has trained 20 Weighted-ELMs as the weak classifier. It can be seen from Table 5 that when the proportion of the positive and the negative samples reaches 1:100, the ELM algorithm and the Weighted-ELM algorithm have lower AUC value while the AUC value of the WELM-Adaboost algorithm is still maintained at more than 0.8. This shows that the proposed WELM-Adaboost algorithm has better performance.

6. Conclusions

This paper firstly applied the advertising company's big data to build the user graph system for the purpose of classifying the advertisement data. The output of this user graph system includes the user's age, gender, and the interest preferences, which are used as the input of the prediction model of CTR. Experiments show that this kind of features has a significant effect on the CTR prediction.

The main contribution of the paper is to propose a WELM-Adaboost algorithm based approach for the CTR prediction of the RTB advertisement. We applied the real advertisement dataset to implement the experiments by applying the AUC value as the measurement criteria. We compared both the ELM algorithm and the Weighted-ELM algorithm with the proposed approach. The experimental

TABLE 6: The Comparison of the AUC value of the three algorithms.

Proportion	ELM	Weighted-ELM	WELM-Adaboost
1:5	0.951	0.882	0.962
1:10	0.838	0.831	0.919
1:20	0.840	0.870	0.932
1:50	0.839	0.840	0.919
1:100	0.682	0.583	0.838
1:150	0.508	0.614	0.789
3:1000	0.503	0.569	0.679

results show that the AUC value of the proposed algorithm is significantly improved compared to the ELM and the Weighted-ELM based method.

Although this paper has made a systematic study on the feature extraction and the CTR prediction of the RTB advertisement, there are still some issues to be improved in the future.

The deep neural network may be a good way for the further future study of the CTR prediction.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

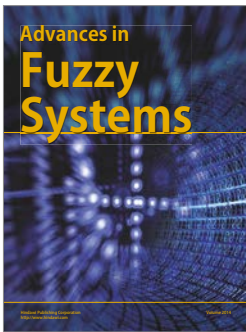
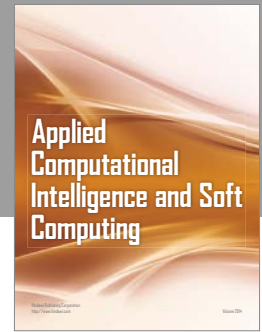
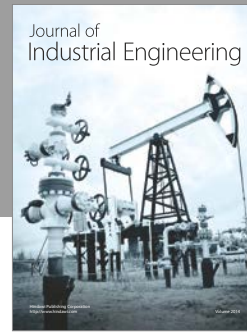
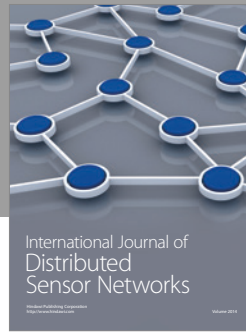
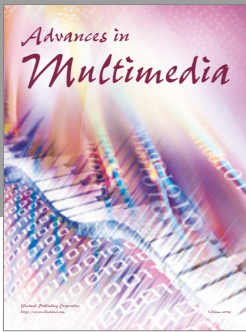
Acknowledgments

This paper is funded by the National Natural Science Foundation of China (nos. 61673056 and 61673055).

References

- [1] Z. Meng, *Research on the Personalized Advertising Push Services for Internet Users*, Donghua University, Shanghai, China, 2014.
- [2] A. K. Menon, K. Chitrapura, S. Garg, D. Agarwal, and N. Kota, "Response prediction using collaborative filtering with hierarchies and side-information," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '11)*, pp. 141–149, San Diego, Calif, USA, August 2011.
- [3] M. Richardson, E. Dominowska, and R. Ragno, "Predicting clicks: estimating the click-through rate for new ads," in *Proceedings of the 16th International World Wide Web Conference (WWW '07)*, pp. 521–530, May 2007.
- [4] O. Chapelle, "Modeling delayed feedback in display advertising," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14)*, pp. 1097–1105, New York, NY, USA, August 2014.
- [5] Di. Shao, *Research on High Level Feature Representation and Predicting Methods in Online Advertising*, Harbin Institute of Technology, Harbin, China, 2014.
- [6] T. Fawcett, "ROC graphs: notes and practical considerations for researchers," *Machine Learning*, vol. 31, no. 1, pp. 1–38, 2004.
- [7] W. Xiao-Shu, *Click-Through Rate Prediction Based on Deep Neural Network Model*, Beijing University Of Posts And Telecommunications, Beijing, China, 2015.

- [8] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489-501, 2006.
- [9] W. Zong, G.-B. Huang, and Y. Chen, "Weighted extreme learning machine for imbalance learning," *Neurocomputing*, vol. 101, pp. 229-242, 2013.
- [10] X. Zhang, Y. Zhuang, W. Wang, and W. Pedrycz, "Transfer boosting with synthetic instances for class imbalanced object recognition," *IEEE Transactions on Cybernetics*, no. 99, pp. 1-14, 2016.
- [11] K. Li, X. Kong, Z. Lu, L. Wenyin, and J. Yin, "Boosting weighted ELM for imbalanced learning," *Neurocomputing*, vol. 128, pp. 15-21, 2014.
- [12] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *Proceedings of the IEEE International Joint Conference on Neural Networks*, vol. 2, pp. 985-990, July 2004.
- [13] H.-J. Rong, Y.-S. Ong, A.-H. Tan, and Z. Zhu, "A fast pruned-extreme learning machine for classification problem," *Neurocomputing*, vol. 72, no. 1-3, pp. 359-366, 2008.
- [14] Y. Xu, Q. Wang, Z. Wei, and S. Ma, "Traffic sign recognition based on weighted ELM and AdaBoost," *IEEE Electronics Letters*, vol. 52, no. 24, pp. 1988-1990, 2016.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

