

Does Ad-Context Matter on the Effectiveness of Online Advertising?

Caihong Sun, Meina Zhang, Meiyun Zuo*
Key Laboratory of Data Engineering and Knowledge Engineering,
School of Information, Renmin University of China
Beijing, China
chsun@ruc.edu.cn zhangmeina@ruc.edu.cn zuomy@ruc.edu.cn

Abstract—In this paper, we focus on examining the effects of Ad-context on the click-through rate (CTR) for the online advertising. Many researches have shown that ad-context congruity is a key factor to CTR, but the features of ad-context are rarely introduced in CTR prediction algorithms. By leveraging data from various sources, and using text mining and sentiment analysis techniques, our proposed approach extracts three types of features (i.e. users, advertisements and ad-context features) to predict CTR. User features describe “Who” is browsing the webpage, advertisement features depict “How” an ad serving is, and ad-context features include “What” the product is, “Where” an ad displays, as well as the “Mood” of the context. Experiment results show that our proposed approach outperforms the benchmark models by introducing ad-context features. Novel ad-context features we proposed make good contributions to the prediction of CTR. The research highlights the power of ad-context in online advertising CTR prediction. Moreover it gives an insight into supporting effective advertising and improving users’ satisfaction.

Keywords—Click-Through-Rate, Ad-Context, Sentiment Analysis, Feature Engineering, Online Advertising

I. INTRODUCTION

With the increasing of Internet users, online advertising becomes a marketing gold mine and generates huge advertising revenues. For example, the Interactive Advertising Bureau (IAB) announced on December 20th 2017 that digital advertising revenues in the US for the first half of 2017 surged to \$40.1 billion, a 23 percent year-over-year rise from the \$32.7 billion reported during the same timeframe in 2016. The huge commercial value of online advertisement in some extent depends on whether users click on the advertisement or not. The click-through rate (CTR), as a major measurement of online advertising, can be used to improve users’ satisfaction and support effective advertising. It plays an important role in online advertising. Academia and industry have also invested a lot of effort in the CTRs prediction model of advertising.

The features used to predict CTR in the existing literatures generally could be classified into three kinds: advertisement features, users’ features and ad-context features. Many researchers use the advertisement features, such as the size, the position of an advertisement, and users’ features such as demography, the equipment type and user’s behavior history for CTR prediction. But few researchers introduce ad-context features into CTR prediction algorithms since they are not easy to be obtained. There are two contrasting theories to explain the relationship between ad-context congruity and memory for ad content, where Ad-context congruity is defined as the degree to which advertising material is thematically similar to adjacent editorial content [1]. Priming theory leads to increased memory of ads that are similar to their context, but interference theory is prone to decrease memory when ads resemble their context. Although contrasting, both priming and interference theories show that ad-context matters in advertisement effectiveness [1-3]. In the study of contextual advertising, Chakrabarti et al. measured the ad-context congruity by using the cosine distance between the keywords of the advertisement and the webpage the ads placed [4]. By adding this information into CTR prediction, it gained 25% prediction accuracy improvement. But what the ad-context features are and how to extract them?

In this paper, our main research question is to extract ad-context features and introduce them into the state-of-art prediction algorithms to enhance the performance of CTR prediction. The main contributions of this study are in two areas:

First, this work demonstrates a prediction framework in which the different sources of data could be collected, fused and analyzed to train CTR prediction algorithms, such as Logistic Regression (LR), Gradient Boosting Decision Tree (GBDT) and Factorization Machine (FM). In this study, we collected a SUV automobile advertisement dataset and crawled the webpages that the ads placed to demonstrate the process and the effectiveness of our prediction framework.

Second, our study proposed several novel ad-context features: “What”, representing the attributes of a product on a web page (i.e. price, the origin of production and product type etc.), “Where”, ad serving channels and the “Mood” of the web contents. Although many features have already been studied and introduced in CTR prediction for better performance in many aspects, the effect of ad-context features is rarely considered into CTR prediction until now, especially the features in product and mood dimensions. We showed that these features make

great contributions to the system's performance and help to explain that the ad-context is a key factor to the effectiveness of online advertising.

Accuracy, logLoss and AUC (the Area under Receiver Operating Characteristics (ROC) Curve), will be employed as the evaluation criteria for the performance in our comparison study.

The paper proceeds as follows. Section 2 reviews the research literatures. Section 3 outlines the system framework and presents the process of feature engineering. Section 4 describes the data collection and shows the experiment studies. Section 5 concludes the paper with future remarks.

II. LITERATURE REVIEW

In this study, we draw mainly on three streams of researches in online advertising: (i) CTR prediction models, (ii) features used in CTR prediction, and (iii) ad-context.

A. CTR Prediction Models

Many researchers and advertisers to predict CTR use logistic Regression. The major difference of these studies is the features they used for prediction. Robinson et al. studied how 7 features of banner ad affect the online advertising effectiveness by using LR, and they found that in network game domain, the size of ad, promotion and game information were good for clicking, but the dynamic banners, the corporate brands or flags had no effect on clicking [5]. Chakrabarti et al. used LR and studied how the congruity between ad and web contents influences the CTR in contextual advertising [4]. Besides the classic LR model, Dave and Varma applied GBDT to estimate CTR [6], while the main idea of GBDT is inspired by probably approximately correct learning model proposed by Valiant. The GBDT algorithms are prone to over-fitting, since they ignore the correlations among features. FM algorithm proposed by Rendle [7] overcomes this drawback, and FM can handle the dependent features well. Zhang et al. used RNN (Recurrent Neural networks) to predict the CTR in search advertisements [8]. RNN outperforms LR when data volume is large. Moreover, to improve the accuracy of CTR prediction, one strategy is to fuse several models together. The simple one is to weighted add the results of each models, or stacking these models. Tian et al. use FM+GBDT, LR+GBDT to evaluate the effectiveness of feature selections [9].

B. Features in CTR Prediction

The features used by CTR prediction could be classified into three kinds: the features of advertisement, the features of users and ad-context features. These three kinds of features constitute the feature triple<advertisement, user, context>. The advertisement features, such as the size, position and dynamics of an advertisement, are often used to predict CTR prediction [5]. The features of users i.e. the demography of users, the equipment type, user behavioral data and user ID, are introduced into CTR prediction models to improve the accuracy of CTR prediction [10]. Ad-context features are rarely used in

literatures since they are not easy to be acquired. Richardson et al. [11] introduce the degree of congruity between the search keywords and an advertisement to estimate the CTR of a new advertisement. Chakrabarti et al. [4] improve the prediction effectiveness with 25% by introducing the content congruity between an advertisement and the web page it placed. In addition, many CTR prediction literatures try to introduce more features to improve the prediction accuracy [12-13]. To sum up, the ad-context features are not well explored in CTR prediction yet.

C. Ad-context

Although ad-context is considered to be important in CTR prediction, to the best of our knowledge, there are few literatures discussed about ad-context features engineering. In the studies of ad-context in online advertisement, many researchers focus on exploring the effects of ad-context congruity which is defined as the degree to which advertising material is similar to adjacent editorial content [1]. The research on the ad-context congruity is mainly focus on the following perspectives: structure e.g. background color [14], mood [16], semantic e.g. whether deliver an automobile ad to the auto, finance or sport website [17]. All these contextual factors will influence the effectiveness of advertising. However, articles that study contextual factors from the web product dimension are basically not available. Especially on some vertical website, the semantic and structural factors are completely the same with those of the target advertising. Since the product features such as price have impact on advertisement effects, the product information on the web page for a vertical advertising could possible affect users' click. Psychological research shows that congruity of the attitude related, has important impact on the individual cognition and attitude formation [14-15]. Mood has impact on the audience's attitude to an advertisement and does influence the effectiveness of an advertisement [16]. Some studies [4, 11] demonstrate that semantic or thematically contents influence the effectiveness of an advertisement a lot. To sum up, ad-context does have impact on advertisement effect. Since ad-context features are rarely discussed from the feature engineering aspects, in this study we try to extract ad-context features from web contents and utilize them to improve the CTR prediction.

III. CTR PREDICTION FRAMEWORK AND FEATURE ENGINEERING

A. CTR Prediction Framework

Figure 1 illustrates the framework of our CTR prediction. The first phase is data acquisition. We picked two complementary data sources—log data from an advertising agency and their ad serving websites. We can acquire the features of ads, the impression and click data and some user features from the log data. And the ad serving website provides the ad-context features such as serving channels, product features on the website and mood of a webpage. As for data collection methods, the two sources are different as well. We use API to extract data

from the log data provided by the advertising agency, while crawl the contents of the website via python.

In the second phase, data from both sources are cleaned, transformed, consolidated, and stored in a database. In particular, we use text mining and sentiment analysis techniques to get the mood feature of the web contents and then stored them in database.

The third phase, “feature engineering”, involves using the acquired data to construct features that will ultimately be used to train a CTR predictive model. Specifically, we classify various features into three groups, which will be described in detail later.

In the fourth phase, a predictive model can be trained. The Experimental Result Analysis section will discuss our experiments and research questions in detail.

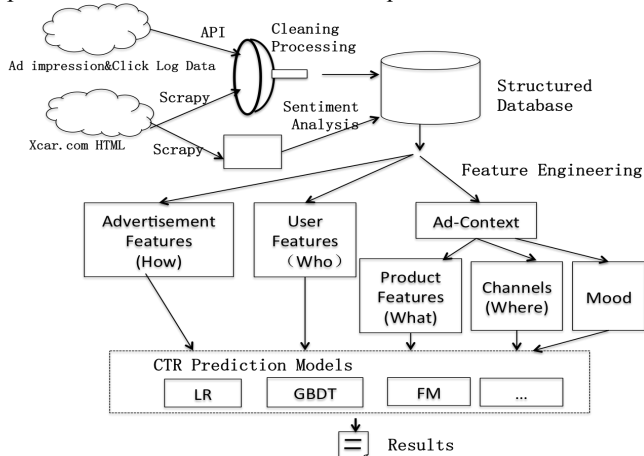


Figure 1. The Framework of CTR Prediction Feature Engineering

B. Feature Engineering

In our study, we derived three groups of features: “who” features, “how” features, and ad-context features. Ad-context features consist of ad serving channels (Where), and product features (What) and mood features.

“Who” features consist of the features describing users who could view the ads: UserID, the number of visited pages, the number of browsed auto types, city, the number of visited channels, and type of the equipment a user used.

“How” features describe how an advertisement displays: AdID, Ad’s size, position, the number of impressions, the number of impression days and so on.

“Ad-context” features consist of ad serving channel (Where), product features (What) and mood feature. The serving channel captures the channel targeting aspect, where the ad was served in a web site. Since our study focuses on serving ads on the vertical media, different channels provide consumers with different types of information, and consumers may browse different channels at different consumption stages. Therefore, the click through rate of ads will also vary in different channels. While product features describe the semantic aspect of web contents by adding the product dimension which the ad described. The difference of product features between the

advertising target and the webpage has impact on the consumers. Here the product features are described as product ID, price, the origin of production and product type since some studies show that the price, the original site, and the automobile type are the three main factors that affect the car purchase decision [18]. Furthermore, the mood feature describes the mood of a webpage via the negative or positive attitude, which captures the emotion of the web contents.

IV. EXPERIMENTAL DESIGN AND ANALYSIS

In this section, we conduct several experiments to validate the effects of ad-context features based on LR, GBDT and FM models. We do the experiments by adding ad-context features or not to demonstrate the effectiveness of our proposed method.

A. Data Description

There are two data sources in our research. One is the log data provided by advertising agency. They provide us the features of advertisements, some feature of users and the log data. The other data is crawled by us from the websites where the ads delivered to access the ad-context features. In our study, we focus on a new SUV auto ads activity. The ad-serving period is two weeks, from December 23rd 2014 to Jan 5th 2015. Xcar.com is the ad-serving website, an auto vertical media where the serving channels include 4 channels: information, shopping guide, test-driving and car usage.

We extract user and advertisement features from the data provide by the advertising agency, and the ad-context features from the web contents we crawled. Serving channel feature (“Where”) has 4 categorical values according to the channels the advertisement displayed on. These four channels provide information consumers they sought for at different stages of their purchase decision. The information channel provides some basic information about cars, which satisfied the information need of consumers in the early stages of purchase decisions. The shopping guide channel provides consumers with specific information on a topic, such as price, auto type, for comparison. Test-driving channel delivers some driving experience of a certain car type. The car usage channel provides maintenance information after purchase. The source of mood factors is from the textual context of the web pages. We apply the dictionary-based sentimental word analysis algorithm to the text content of the webpages we crawled. In our experiments, we adopt the sentimental lexicon from the information retrieval laboratory of Dalian University of Technology for sentimental analysis. The mood feature is a numeric type, while a negative or positive value represents negative or positive mood respectively. The product features (“What”) consist of price, the origin of production and product type.

The price of SUV, the target product of advertising activity, in the ad activity is about 129.9-169.9 thousand RMB. The total impression of the ad is 440812, the number of clicks is 586. There are 251199 users, and 501 users click the ads, as shown in Table 1.

TABLE I. STATISTICS OF THE DATASETS

Number of Impression	440812
Number of Clicks	586
Number of users	251199
Number of clicked users	501

B. Evaluation Metrics

For CTR prediction, we tried 3 algorithms: LR, GBDT and FM. We evaluate the overall performance based on 3 metrics:

1) *Classification accuracy*, which is the percentage of correctly predicted instances;

2) *Logloss*, a measure defined as the negative log-likelihood of the true labels given a probabilistic classifier’s predictions, the smaller the value is, the better the prediction;

3) *AUC*, the curve plots the true positive rate against the false positive rate. An AUC of 1 means a perfect classification whereas 0.5 refers to a random guess. Being more robust against prior distributions, AUC is considered by many researchers to be one of the best indicators of a classifier’s performance.

C. Three State-of-art CTR Prediction Algorithms

To illustrate the effectiveness of our proposed method, we choose the following three state-of-the-art CTR prediction algorithms for comparisons.

- **LR** (Logistic Regression): It is a widely used method for CTR prediction, especially used as the baseline model for comparisons with other prediction methods [4][6].
- **GBDT** (Gradient Boosting Decision Tree): It is a non-linear model, an ensemble decision tree inspired by boosting. It has the advantages to discover the raw features and cross features. GBDT usually used in CTR prediction fused with other methods such as LR and FM [9][18].
- **FM** (Factorization Machine): Factorization machines (FMs) are a generic approach that allows mimicking most factorization models by feature engineering [7]. FM is able to perform a regression model whose task is to estimate a function $y: \mathbb{R}^n \rightarrow \mathbb{T}$ from a real valued vector $x \in \mathbb{R}^n$ to a target domain \mathbb{T} and it could model all single and pairwise interactions between the input variables by using factorized interaction parameters.

D. Experimental Design and Analysis

To evaluate the effectiveness of ad-context features, we apply the features with and without ad-context features to three algorithms respectively: LR, GBDT and FM. Since the click rate is very low, i.e. the positive and negative samples are extremely misbalanced. In our experiments, we sample the ratio of positive and negative from 1:1 to 1:9, and we found the sampling ratio of positive and negative samples mattered just a little bit in all three algorithms, When the ratio is 1:5, the performance is better. As shown in Figure 2, with the ratio of sampling in positive and negative data from 1:1 to 1:9, the performance of LR varied a little (we can find the similar results in both GBDT and FM). In the following experiments, we set

the positive and negative sampling ratio as 1:5.

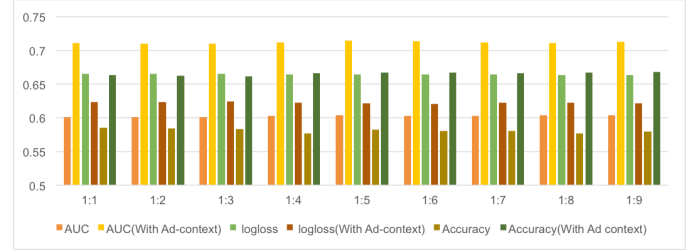


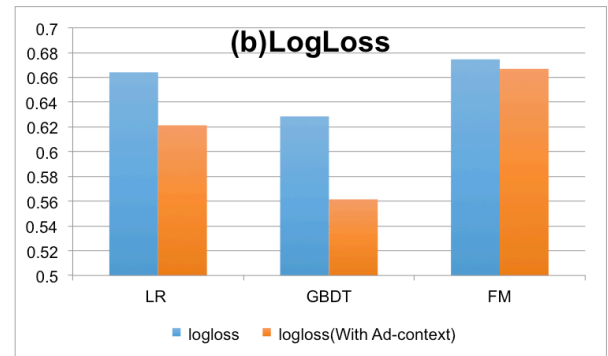
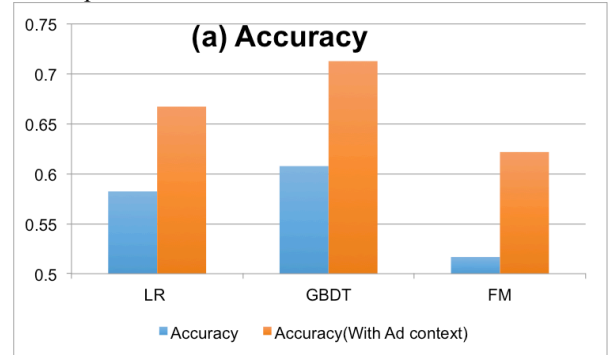
Figure 2. Performance with and without ad-context features by using LR in different sampling ratio

Experiments are designed to answer the following two research questions.

RQ1: Does Ad-context matter in CTR prediction of online advertising?

We design comparison experiments by adding ad-context features or not. The baseline is to use features of advertisement and users to predict CTR, without any ad-context features.

As shown in Figure 3, by introducing ad-context features, all the three evaluation metrics get better in LR, GBDT and FM. The accuracy improves a lot (shown in Figure 3 (a)), the logloss decreases (shown in Figure 3(b)) and the AUC increases (shown in Figure 3(c)) in all three algorithms. In our experiments, GBDT outperforms LR and FM.



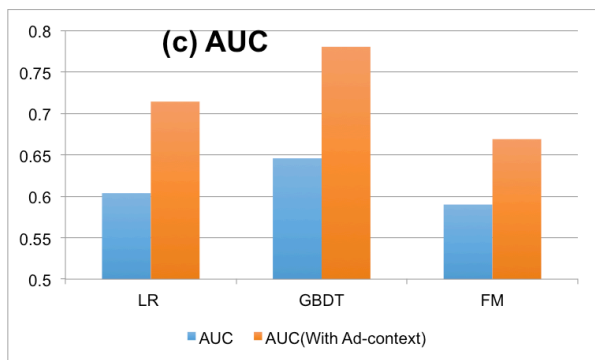


Figure 3. Comparisons in (a) Accuracy (b)LogLoss and (c) AUC with and without ad-context features

To sum up, ad-context features do matter in the prediction of CTR in our experiments.

RQ2: Are all dimensions of ad-context we proposed valuable?

To illustrate the effect of “Where”, “What” and “Mood” dimensions in ad-context, we design experiments by adding each of them into baseline in LR algorithm, and we find (shown in Figure 4) that “Where” and “What” features improve all three evaluation metrics individually, but the “Mood” feature does not. Furthermore, by introducing all three kinds of ad-context (Where, What and Mood), the algorithm gets better performance.

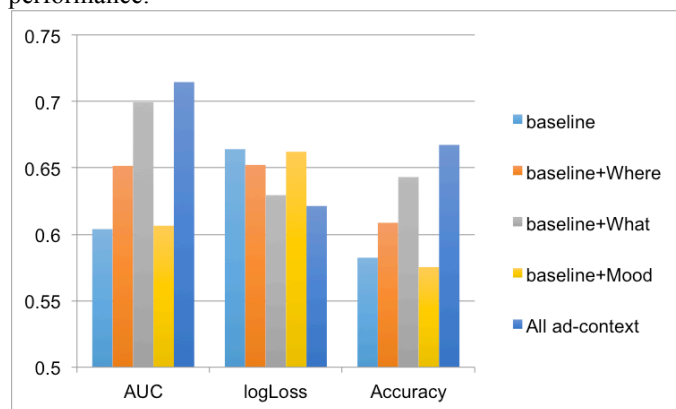


Figure 4. The Effects of “Where”, “What” and “Mood” Features in LR algorithms

As shown in Figure 4, we can see that “What” features, i.e. the features of products in our experiments, contribute a lot in CTR prediction. By adding all three kinds of ad-context features, the algorithm gets much better performance. The similar findings are found in GBDT and FM. The experiments demonstrate that our proposed ad-context features are very effective on CTR prediction.

V. CONCLUSIONS AND FUTURE WORK

In this study, we proposed a CTR prediction framework by collecting data from different sources. By employing text mining and sentiment analyzing, we extract ad-context features including “Where” the ads displayed, “What” is the product and

“mood” of the web contents. CTR prediction improves a lot with the novel ad context features we proposed in LR, GBDT and FM algorithms. Different from the application of context factors in search advertising mentioned in the previous literature review, this article attempts to apply context factors in general banner ads. Existing studies aim to explore how ad-context congruity affects the effective of online advertising. Here, we are not considering congruity but introducing context attributes to the CTR prediction algorithm. Our study demonstrates that ad-context is valuable to CTR prediction and gives out a practical way of ad-context feature engineering. Our work gives an insight on ad-serving channels decision-making and users’ satisfaction improvement. There are also several directions for future research. Our future work includes testing our method on more datasets and CTR prediction algorithms. More ad-context features could be extracted and drawn into further study.

ACKNOWLEDGEMENT

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 10XNJ065, part by the National Natural Science Foundation of China under Grant 71771210, part by the Beijing Natural Science Foundation under Grant 9182008, part by National Social Science Foundation of China Major Program under Grant 13&ZD184. Meiyun Zuo is the corresponding author of this paper.

REFERENCES

- [1] Gunter, B., et al., Children’s memory for television advertising: effects of programme–advertisement congruency. *Applied Cognitive Psychology*, 2010. 16(2): p. 171-190.
- [2] Kamins, M.A., L.J. Marks, and D. Skinner, Television Commercial Evaluation in the Context of Program Induced Mood: Congruency versus Consistency Effects. *Journal of Advertising*, 1991. 20(2): p. 1-14.
- [3] Yinon, Y. and M.O. Landau, On the reinforcing value of helping behavior in a positive mood. *Motivation & Emotion*, 1987. 11(1): p. 83-93.
- [4] Chakrabarti, D., D. Agarwal, and V. Josifovski. Contextual advertising by combining relevance with click feedback. in *International Conference on World Wide Web*. 2008.
- [5] Robinson, H.R., A. Wysocka, and C. Hand, Internet advertising effectiveness: The effect of design on click-through rates for banner ads. *International Journal of Advertising*, 2007. 26(4): p. 527-541.
- [6] Dave K S, Varma V. Learning the click-through rate for rare/new ads from similar ads[C]// *International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2010:897-898.
- [7] Rendle, S. Factorization Machines. in *IEEE International Conference on Data Mining*. 2010.
- [8] Zhang, Y., et al., Sequential Click Prediction for Sponsored Search with Recurrent Neural Networks. 2014: p. 1369-1375.
- [9] Tian, C., et al., Research and implementation of feature extraction methods of Internet CTR prediction model. *Computer Application Research*, 2017. 34(2): p. 334-338. (in Chinese)
- [10] Zhu, Z.A., et al., A novel click model and its applications to online advertising. 2010: p. 321-330.
- [11] Richardson, M., E. Dominowska, and R. Ragno. Predicting clicks:estimating the click-through rate for new ads. in *International Conference on World Wide Web*. 2007.
- [12] Ta, A.P. Factorization machines with follow-the-regularized-leader for CTR prediction in display advertising. in *IEEE International Conference on Big Data*. 2015.

- [13] Zhang, S., Q. Fu, and W. Xiao, Advertisement Click-Through Rate Prediction Based on the Weighted-ELM and Adaboost Algorithm. 2017. 2017(1): 1-8.
- [14] Chaiken, S., E.M. Pomerantz, and R. Giner-Sorolla, Structural consistency and attitude strength. *Attitude Strength Antecedents & Consequences*, 1995.
- [15] Berkowitz, L., Mood, Self-Awareness, and Willingness to Help. *Journal of Personality & Social Psychology*, 1987. 52(4): p. 721-729.
- [16] Faseur, T. and M. Geuens, Different Positive Feelings Leading to Different Ad Evaluations: The Case of Coziness, Excitement, and Romance. *Journal of Advertising*, 2006. 35(4): p. 129-142.
- [17] Gunter, B., et al., *Children's memory for television advertising: effects of programme-advertisement congruency*. *Applied Cognitive Psychology*, 2010. 16(2): p. 171-190.
- [18] Engel, J.F., R.D. Blackwell, and P.W. Miniard, *Consumer Behavior (8th Eds.)*. 1995.
- [19] Chapelle O, Manavoglu E, Rosales R. Simple and scalable response prediction for display advertising [J]. *ACM Transactions on Intelligent Systems & Technology*, 2013, 5 (4): 1-3