Journal of Computers Vol. 31 No. 3, 2020, pp. 256-265
doi:10.3966/199115992020063103020

The Study on Preventing Click Fraud in Internet Advertising

Zhi Li[1], Weichen Jia[2*]

[1] School of Media and Law, Zhejiang University Ningbo Institute of Technology, Ningbo, China
lizhi@nit.zju.edu.cn

[2] School of Education, City University of Macau, Macau, China
jwc19890114@163.com

Received 3 June 2019; Revised 8 August 2019; Accepted 6 January 2020

Abstract. Since the 21st century, China's internet technology has continuously developed in stride. It has evolved from a convenient and efficient information transmission tool to a diversified information carrier with large number of users and a network social service application platform, having positive impact on people's life. This study analyzes the current strategies to prevent click fraud, and the implementation process of all strategies is basically the same: firstly, customers are threatened by click fraud, and then they adopt rights protection behaviors, adopt different strategies, investigate the causes of click fraud, and compensate for the losses caused by click fraud. The strategy to prevent click fraud proposed in this paper starts from the effective number of clicks. Even if there is a malicious click, the click behavior will be regarded as invalid, and no record will be made in the database of the service provider, so that customers will not be threatened by click fraud. This paper uses the random forest algorithm in machine learning to classify features to determine fraudulent click behavior. The results show that the proposed method is higher than 91% in the prediction accuracy of positive and negative samples. In addition, in comparison with several other classification methods, the random forest classification algorithm Internet advertising click fraud detection effect is the most effective. The results of CPC advertising show that the algorithm is effective and feasible.

Keywords: click fraud, Cost-per-Click (CPC), Internet advertising, strategy

1 Introduction

Presently, the existence of illegal profit and unfair competition in the internet advertising market has become the main reason for click fraud. Common internet advertising billing patterns are mainly divided into three categories. The first category, is according to the advertising's display billing, namely, cost per thousand impressions (CPM), and, cost per targeted thousand impressions (CPTM). The second category, is according to billing by advertising actions, that is, cost-per-click (CPC), pay-per-click (PPC), cost-per-action (CPA), and cost for per lead (CPL). The third category: is according to the sales of advertising, specifically, cost-per-order (CPO), cost for per sale (CPS), and pay-per-sale (PPS) [1]. The most common advertising billing methods are CPM and CPC. Sina, NetEase, Sohu, Tencent and other banner advertising portals use CPM, while Google, Baidu, Qihoo 360, and other search engine sites use the CPC billing advertising method. Internet click fraud has become a significant problem that threatens the online advertising market, some enterprises have taken various measures, such as fraudulent clicks, in order to rank their advertisements or crack down on competitors in the same industry. Consuming competitor's advertising investment costs to achieve their own interest [2].

Compared to other methods, malicious competition and click fraud phenomenon is more obvious in CPC advertising. However, its cause of click fraud is also more complex, a possible reason for this fraudulent action may be the company or website wants to gain an edge over their competitors. Taking the CPC billing advertising method as an example, this study provides in-depth study of fraud click

* Corresponding Author

issues, and puts forth prevention and detection strategies that are beneficial to maintaining the order of the internet advertising market, sustaining the healthy development of network service platforms, safeguarding the rights of advertisers and the healthy development of the internet advertising industry. At the same time, by analyzing internet technologies, this study proposes effective measures to restrict fraud, aiming to effectively create a healthy network environment, control the occurrence of fraud, and avoid losses to the internet advertising industry.

Malicious click in China has begun to flood, and currently out of control, which will greatly threaten the development of Chinese search market. If left unchecked, the consequences could be severe. Click fraud has a more and more serious impact on the profit model bidding ranking at the core of search engine advertising market, when advertisers find that click rate cannot bring corresponding business opportunities. At the same time, they have to pay a lot of click fees, they will reduce the amount of online advertising, seriously damage the development of online advertising industry, search engine service providers put forward a serious challenge! Once "click fraud", the effectiveness of online advertising will be significantly lower, not only to advertisers, especially those advertisers pay high click prices) have very serious consequences, bound to lead to advertisers to gradually reduce the investment cost of online advertising, and even give up Internet advertising, but also makes the Internet advertising service providers are worried about their future prospects for development.

From the technical point of view, click fraud should not only be controlled from the technical point of view, but also involve the issue of integrity. If you want to cheat, no technology can be completely overcome and directly impact its market future. To solve click fraud well, or develop an effective strategy to prevent click fraud, or replace the current bidding ranking with another new business model, but these cannot be realized in a short time, the current legislation is not perfect, we must consider the use of laws to strengthen the constraint on click fraud. Therefore, the purpose of this study is to prevent the rampant click fraud, reduce the harm of click fraud to the Internet, and make efforts to build a healthy and harmonious network.

2 Related Work

2.1 E-Commerce

In the field of e-commerce, advertising helps to increase the popularity of their products and allow more people to buy them. However, with the proliferation of malicious competition in pay-per-click in e-commerce, the phenomenon of click fraud is more obvious. Some companies adopt a series of methods for their own advertising rankings or competitors who are competing with the company, and even use fraudulent clicks in the pay-per-click model to consume the other party's advertising investment costs to achieve their goals. The Google fraud click was a sensation, and a settlement was reached in March 2006 [3]. The result was Google's compensation. At the same time, large search engine companies have established relevant "anti-fraud alliances" to address click fraud. However, the results of cracking down on fraudulent clicks are not very prominent. In April 2008, Yahoo was also accused by e-commerce complaints of fraudulent clicks in the pay-per-click model, saying that Yahoo's indulgence of fraud led to the loss of merchants' interests [4]. Display advertising is an important part of online advertising, which accounts for 40% of the entire online advertising industry. According to eMarketer, in 2017, Facebook and Twitter accounted for 33% of the display advertising market share, and Google's display advertising covered 80% of the world's Internet users.

2.2 CPC fraud

According to analysis of the principle and form of fraud, the click fraud in CPC billing advertisement can be divided as: (1) The internet service platform's partners click to create profit; (2) Malicious clicks by main competitors of internet advertisements, which are mainly used to combat web page rankings or consume advertising budgets to improve their own ad ranking; (3) Inadvertently invalid clicks; (4) Fraud in the internet advertising service platform, to raise and generate advertising, income by hiring cheap labor to click on the advertisement; (5) Internet advertisers' fraud, as the network platform itself has a relatively low ranking in the search engine, they can increase the frequency and exposure through, keyword ranking; (6) Indirect click fraud, the advertisers use false words or images in search engines or

portal pages to increase their exposure; (7) Fraudulent attacks between search engines platforms, vicious competition mechanism makes the other party lose the trust of advertisers and loses their market.

Through the analysis of the fraud process, the types of fraud methods are mainly divided into general fraud methods and complex fraud methods [5].

In the general fraud method, supposing U is a fraudulent user with suspicious behaviors, R is a service platform provider for publishing internet advertising, and T is an advertiser. The steps for click fraud methods of general internet CPC billing ads include:

①When U (suspicious fraudulent user) clicks on the page request from R (internet advertising service platform), the advertising service provider will return page R.html to U, and programmatically send it to T (advertiser), then provides link to U;

②U (suspicious fraudulent user) clicks on T's (advertiser) hyperlink in page R.html;

③T (advertiser) sends the feedback page T.html to U (suspicious fraudulent user);

By this way, T (advertiser) publishes advertisements for R (internet advertising service platform) by click fraud of suspicious malicious U (suspicious fraudulent users), due to the existence of click fraud, payment will be quickly consumed, and causing the network advertising to advance down the line [6].

In the complex fraud method, supposing U is a fraudulent user with suspicious behavior, R is a service provider platform for publishing internet advertisements; S is the internet advertising source site (R advertises on S), and T is the advertiser. Steps to fraudulent clicks on a complex internet CPC billing ads includes:

①When U (suspicious fraudulent user) sends a click request to S (internet advertising source site), S will send a page S.html to U, which feds R's (internet advertising service platform) information to the user;

②When the user opens page S.html, the script contained in it will automatically be sent to R (internet advertising service platform) without the user's click;

③R (internet advertising service platform) feedback page R.html to the user and points to T (advertiser), T (advertiser) then feedback page T.html to the user.

④S (internet advertising source site) uses page R. html to automatically feedback information to U (suspicious fraudulent user), while U will point its information content to T (advertiser), and then feedback the information to the user through page R.html.

2.3 AI/Machine Learning

Artificial intelligence and machine learning have been used in various fields in recent years, and have been widely used to solve complex problems in engineering applications and sciences, and have achieved unexpected results. Artificial neural networks are imitations of biological neural networks that learn by stimulating and forming new connections between neurons [7]. The artificial neural network models the complex relationship between input and output by using a large number of interconnected neuron nodes that have been fixed and using the empirical data to propagate in the forward direction to update the connection strength between the neurons. It is essentially a mathematical model that can be processed by a computer. Zhang Zhiqiang et al. proposed a feature-based learning click rate prediction technology for the click rate problem in search advertising [8]. The method realizes feature dimensionality reduction based on tensor decomposition, and makes full use of machine learning technology to describe nonlinear correlation in data to solve the feature learning problem of high-dimensional sparse advertising data. Liu Mengjuan et al. proposed an online advertising click-through rate prediction model based on fusion structure. The model can flexibly integrate deep neural networks with different structures to learn the high-order representation of the original high-dimensional sparse features, so that the click-rate prediction model can utilize richer high-order feature information [9]. Machine learning is a discipline in which computers build models based on data and use models to simulate human intelligence activities. The biggest advantage of machine learning is that it has generalization ability and can play a huge advantage in the detection of advertising fraud.

3 The Strategy of Preventing Click Fraud in Internet CPC Billing Advertisement

Compared to traditional advertising methods, internet CPC billing advertising can minimize unnecessary waste of advertising resources, so it has been widely recognized by many advertisers, becoming a core and popular way of advertising billing in the internet advertising industry. However, because of the CPC billing advertising's inherent characteristics, it has a strong impact on the internet advertising industry, as well as the large cost of advertising for the advertiser. The advertisers associated with the internet advertising industry, the advertising service provider platform, and the advertising search engines are all faced with suspicious behaviour fraud for the purpose of profit. As a result, it has attracted attention from the industry and academia. We aim to technically and institutionally find strategies to prevent click fraud proposed the following specific internet advertising strategies to prevent click fraud.

3.1 Establishment of an Internet Advertising Service Platform Defense System

First, to use the search engine service platform to track the defense system [10]. The search engine service platform can establish a tracking system to track all cookies and IP independent click data, analyze the daily clicks, the average time of browsing, the web-site per click, the average turnover rate of each click's keyword data, and specific abnormal search engine keywords. It can also monitor hourly, and daily click trends, and the average click through rate data for a specific location [11]. To check whether there is an abnormal phenomenon, indicators include the peak value in clicks at a certain time, the click rate rises with a zero conversion rate, zero page views, and whether previous competitors have a dropped out of the front ranking, including whether one of them has been able to maintain a higher position. By tracing the connection between each keywords and specific IP address, the source of the click fraud, its frequency of, and the resulting economic loss are identified. Although the use of this precaution system can identify the source of click fraud, and recover the amount of loss, form the technical point of view, it is difficult to distinguish whether it is fraudulent clicks or it is user's accidental click. At the same time, a large number of paid keywords and web advertisements need to spend the energy at observing and analyzing, the search engine service platform for enterprises constitute a great deal of work, and the accuracy of the analysis results is also greatly discounted, this has become a major shortcoming of the search engine service platform tracking system [12].

Second. directly using Google, Yahoo, and Baidu search engine's defense systems. At the present, in the internet advertising market, Google provides the best and most advanced search engine for preventing click fraud technology, such as the Google search engine [13]. Google performs statistical analysis on click traffic data, and then deducts suspicious click fraud data from the total number of statistics [14]. When the advertiser chooses the CPC billing advertisement method, Google may assign aspecific resource locator account (URL), and use the log analyzer program to investigate all the click data received by each URL (including date, time, source URL, web page browsing, etc.). After careful observation and analysis, basic determination of whether there is any problem flow can be achieved [13].

Google's search engine is divided into detection, advanced monitoring, filtering, and control to prevent advertisement click fraud:

Step 1: The system will perform click tracking check on advertisement words every time;

Step 2: Monitoing all of data points for each click, including IP address, click time, repeated clicks, and other click patterns;

Step 3: To analyze the relevant factors, isolate and filter potential invalid clicks, and avoid appearing in the URL report.

The Baidu search engine uses automatic filtering, automatic adjustment, or human-computer interaction to prevent advertisement click fraud:

Step 1: Any clicks obtained through the Baidu search promotion must pass the filtering system before billing. According to clicks on the IP, cookies, keywords and other parameters from several dimensions for analysis, if a click exception is found, it will be automatically identified as an invalid click, which will be filtered and not billed.

Step 2: The system automatically analyzes the characteristics of clicks, adjusts the filtering parameters of the online filtering system, and effectively intercepts invalid clicks.

Step3: It automatically checks and analyzes the clicks, and manually analyzes the abnormal clicks to confirm whether they are abnormal. Then feedback the results to the system and automatically adjust the

filter parameters to filter the invalid clicks.

In brief, the Google and Baidu search engines take precautions against malicious click fraud, it is essentially based on tracking detection and filtering.

Third. to use word and picture symbols to verify code guard system. In the process of preventing click fraud, the verification code prevention system is an optimization and improvement compared to the prevention systems of Google, Yahoo and Baidu search engines, it strengthens the accuracy of the click flow analyzer.

When the user clicks on a website advertisement or a keyword advertisement of the search engine bid ranking, the website automatically pops up a window containing the graphic verification code, allowing the user to recognize the input for verification [15].

If the verification is successful, the target page of the advertisement is shown, otherwise it prompts an incorrect notice of the verification code. Finally, on the server side, the click stream data's analysis is performed on the display of the target advertisement page, and an effective click rate is calculated [16].

The word and picture symbol verification code defense system can not only effectively prevent click frauds like Trojan software programs, but also shield the visitors from unintentional invalid clicks. It also greatly reduces manual intentional click fraud, and facilitates the implementation of click stream data base's analysis. However, the purpose of advertisers choosing CPC billing advertisements is to reduce non-target users and attract more potential users. Although the system has obvious advantages, it hangs on the user's patience to browse web advertisements. Which would be accompanies by costs that internet service platforms and advertisers may not be willing to accept.

3.2 The Internet Advertising Service Platform and Advertisers Introduce Third-party Detection to Prevent Click Fraud

In April of 2005, advertisers jointly accused Google of alleged click fraud. In August of 2006, nearly 10 of Baidu's competitors gathered at the Baidu office to protest the auctioning of ranked customers. In recent years, more and more advertising service platforms and advertisers have faced the landmark click fraud events of Google and Baidu. They have chosen third-party network advertising monitoring software to monitor advertisements, and to understand the actual clicks situation. In 2008, Yahoo became the first search engine to cooperate with third parties to prevent click fraud. It committed to preventing the occurrence of click fraud and creating better investment returns for advertisers. Third-party checks on process is shown in Fig. 1:

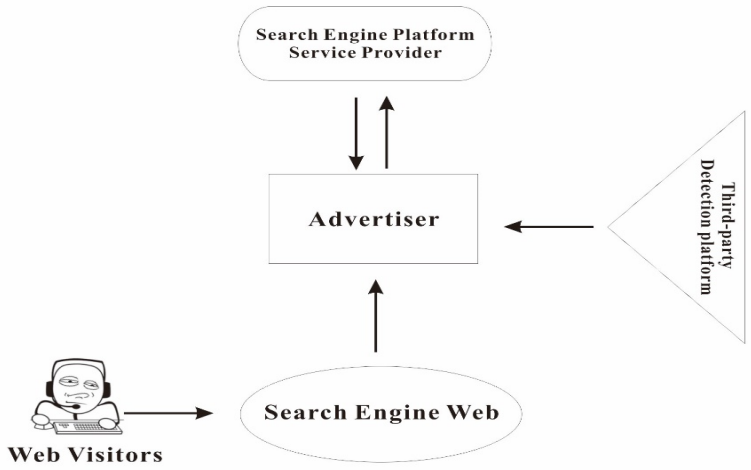

Fig. 1. Third-party detection process

The introduction of third-party detection technologies to prevent click fraud is mainly based on technical measures. The internet advertising service platform and advertisers can set a two-way monitoring program, where a sudden increase in the number of clicks on the advertiser's website, repeated clicks, or abnormal clicks from the same place are constantly traced and monitored.

If the advertiser suspects that the website is victimized by fraudulent clicks, it should quickly contact the search engine platform service provider to report possible fraudulent acts, provide data for feedback,

and continue to monitor the clicks of the advertisements to reduce the advertiser's economical loss. In addition, the search engine platform service provider will also receive third-party monitoring data, which can automatically detect, shield, and filter click fraud, to ensure that the target users are expanded and that it maintains the return efficiency of the advertisers.

3.3 Click Fraud Detection Method Based on Integrated Feature Selection

Random forest classification algorithm. Random forest is an important direction in the field of machine learning. It is robust to data with noise and missing values and has a fast learning speed. Therefore, it is widely used in various classification, prediction and abnormal point detection problems. Here, the random forest is used to classify the data samples. The specific algorithm process is as follows:

The input of the algorithm is the training set T, the number of categories C; the output is the classification result of the random forest classifier model. First, use the boostrap method to select the training set of size N for each tree. Select k features randomly at the node, compare and select the best features, and divide the data set; recursively generate the decision tree without cutting Branch operation. Then calculate the probability that the unknown sample x is classified as C according to the equation $P(C \mid x) = (1/NTree)\sum h_j(C \mid x)$. The majority vote is used to determine the category $C \leftarrow argmax\ P\left(C \mid x\right)$ and the classification error is calculated. Finally, the classification result of the random forest classifier model is output.

4 Fraud Detection Platform Design and Effectiveness Test

4.1 Experimental Design and Data Processing

This paper uses a series of experiments on real online advertising click data to verify the effectiveness of the integrated feature selection detection method proposed in this paper. The experimental environment: operating system windows 10, programming language is python. The dataset used in this paper is a mobile advertising company's real-click fraud detection standard dataset provided by Singapore Management University (SMU) in the FDMA2012 competition organized in 2012. The dataset consists of a publisher dataset and a click dataset (provided in CSV file format) to detect fraudulent publishers who implement illegal clicks from normal publishers. The statistical data used in the experiment is shown in Table 1. The ratio of fraud publishers to normal publishers in the dataset is 5% and 95%, respectively, and has a strong non-equilibrium statistical distribution.

Table 1. Experimental data statistics

Click dataset	Ad publisher dataset		
Number of clicks	Frand	OK	sum
5772649	305(5%)	5776(95%)	6081

The resulting click-to-advertising raw dataset is some basic information about clickers and publishers. The publisher's dataset primarily describes the publisher's information such as the publisher ID (publisherid), address, bank account, and its status (OK). Or Fraud). The click dataset mainly includes numercip (the IP of the clicker in the online user), deviceua (the user agent used by the clicker), publisherid (the unique identifier of the publisher, the link between the two data sets), Campaignid (the unique identifier of the published web ad), referredual (the URL when the ad was clicked), and some characteristic information such as the click time. The two data sets use the publisherid to establish contact, but these features are not available for direct calculation. The raw data is first sorted out. Then, by analyzing the initial features in the original click data set, 118 different predicted feature values and feature-tag pairs corresponding to each publisher are calculated and used for the input of the classifier. In the experiment, 70% of the data in the pre-processed data set was used as the training set, and the remaining 30% was used as the test set.

4.2 Performance Evaluation and Discussion

To measure the classification performance of unbalanced data, a confusion matrix is usually used as a performance evaluation indicator. The common confusion matrix structure is shown in Table 2.

Table 2. Classification prediction confusion matrix

Actual class	Forecast class	
	Fraud	OK
Fraud	TP	FN
OK	FP	TN

According to the confusion matrix, multiple indicators can be used to evaluate system classification performance, such as prediction accuracy (Accuracy), positive class correct rate (acc^+), negative class correct rate (acc^-), G_mean value, and P-R curve. Among them:

$$Accurary = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$acc^+ = \frac{TP}{TP + FN} \tag{2}$$

$$acc^- = \frac{TN}{FP + TN} \tag{3}$$

$$G\_means = \sqrt{acc^+ \times acc^-} \tag{4}$$

The experimental data is extremely unbalanced and requires unbalanced processing of the original data. The method used in this paper is to oversample the positive samples and undersample the negative samples. This method can avoid the loss of important information caused by under-sampling and over-fitting caused by over-sampling to some extent. The experimental results are shown in Table 3. After the data balance processing of the sampling method, the prediction accuracy of the positive class using the random forest classifier is more than 93%, which has a good target effect. The prediction accuracy rate for negative classes has also reached more than 91%, and the overall effect is better.

Table 3. Random forest experiment results

Experiment number	Accuracy	acc+	acc^-	G_means
1	0.9366	0.9324	0.9215	0.9037
2	0.9312	0.9516	0.9106	0.9261
3	0.9207	0.9311	0.9162	0.9158
4	0.9125	0.9469	0.9413	0.9502

To further illustrate the effectiveness of the proposed method, the random forest algorithm proposed in this paper is compared with support vector machine (SVM), Naive Bayes, and decision tree algorithm. Fig. 2 shows the P-R curves of these different classification algorithms. The abscissa and ordinate of the P-R curve are the recall rate and the precision. The closer the equilibrium point of the P-R curve is to the coordinates (0, 1), the better the performance of the classifier. Analysis Fig. 2 found that the random forest classification algorithm based on integrated feature selection proposed in this paper is the most effective in online advertising click fraud detection.

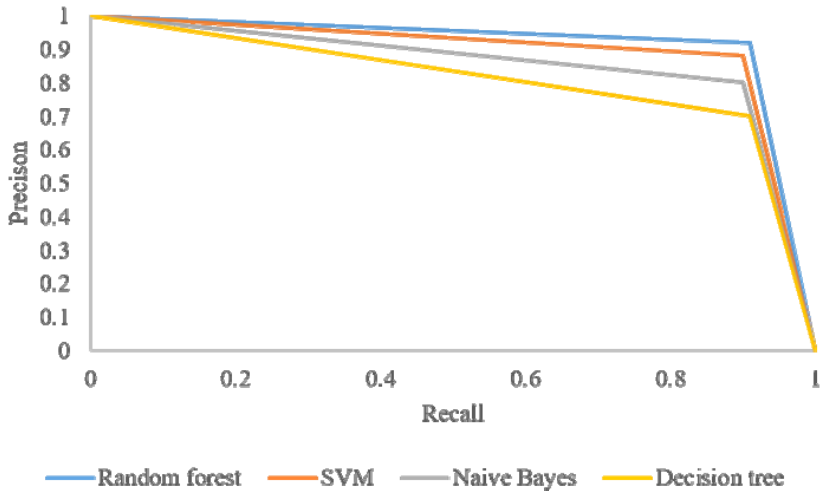

Fig. 2. P-R curves of different classification algorithms

5 Conclusions

With the continuous development of the Internet in China today, click fraud in the internet advertising market is becoming more and more rampant. Although the "People's Republic of China Network Security Law", "Advertising Law of the People's Republic of China", "Law of the People's Republic of China for Countering Unfair Competition" and "Law of the People's Republic of China on Protecting Consumer's Rights and interests" have been successively implemented, the laws and regulations in the form of internet advertising have not been introduced. There are no specific laws and regulations that restrict click fraud and other issues in internet advertising, and the click fraud phenomenon is still difficult to prevent [17].

First, there is no standard measure and precise definition of click fraud. In the internet advertising market, it is difficult to define standards of click fraud, unintentional clicks, and damage to network users or advertisers, and it is difficult to define responsibility. In particular, investigations and evidence collection involve many legal issues. It has brought great difficulties to filing and trial work. Second, internet advertising has not been incorporated into the national advertising management system. Click frauds are not entitled to seek legal solutions. If you want to solve the clicking fraud phenomenon, either technically develop a system that can effectively prevents click fraud, or use another new business model to replace the current bidding rankings, but neither can be achieved quickly in the short term. At the present, in the face of the incompleteness of the national internet legislation, the best way to restrict click fraud is through industry self-regulation.

In view of the fact that the development of internet advertising is still in its immature stage, it has been developing rapidly. It has provided powerful and influential business opportunities for internet advertising service platform. The main goal of internet advertising is to effectively tap potential users. CPC billing advertisements have undoubtedly become the core profit model of search engines. Click fraud has derived from it, and the trouble encountered are unprecedented. Malicious click fraud is a major hidden danger in internet advertisements. Too many malicious clicks can cause great increase in advertising costs, which affects advertisers' trust in internet advertising service platforms. How to prevent or eliminate the impact of click fraud and resolve the trust mechanism between advertisers and search engine providers will become an important direction in the future [13].

5.1 Limitation

The use of the "internet advertising precision delivery platform" in the field of "CPC" mainly encountered problems in data analysis. In this study, we intend to use real dataset to establish a simulation environment. However, in practice needed, we found that exists some difficult. We hope that use the next generation technology based on the real dataset and deep neural network to generate a

simulation environment. And the researcher could observe the platform's changing when they modify some exact parameters.

5.2 Conclusions

This study proposed improve prevention strategies and the establishment of an accurate advertising platform can contribute to click fraud of online advertisements, and that legislation and law enforcement of online advertisements should be strengthened to restrain click fraud, so as to jointly create a healthy and honest online environment. At the same time, it is more desirable for countries to constantly improve relevant laws and regulations from the perspective of the legal system. Internet advertising service platform is constantly improving from the perspective of technology to prevent click fraud, so the internet advertising can grow in a healthy and orderly environment and becomes a new subject in the future of advertising industry.

Reference

[1] L. Zhang, Y. Guan, Detecting click fraud in pay-per-click streams of online advertising networks, IEEE Computer 17(6)(2008) 77-84.

[2] J. Wang, L. Zheng, Analysis of click fraud in Internet advertising, Journal of Management &Technology of SME 23(2008) 59-60.

[3] X. Wang, S. Ding, Click fraud in online advertising and its countermeasures, China Market Supervision Research 4(2007) 15-18.

[4] G. Xie, X. Shi, Reasonable definition of fault liability of network search service providers— Re-evaluation of the judgment of "Yahoo Case" and "Baidu Case", Intellectual Property 18(1)(2008) 81-86.

[5] W. Gao, L. Wang, R. Jin, One-pass AUC optimization, Computer Science 2(36)(2013) 906-914.

[6] O. Chapelle, E. Manavoglu, R. Rosales, Simple and scalable response prediction for display advertising, ACM Transactions on Intelligent Systems & Technology 5(4)(2015) 1-34.

[7] W. Zhang, L. Guo, P. Zhai, Y. Chen, Dynamic characteristics of small world neural networks based on synaptic plasticity. Journal of Biomedical Engineering 35(4)(2018) 15-23.

[8] Z. Zhang, Y. Zhou, X. Xie, P. Pan, Research on ad click rate prediction technology based on feature learning, Chinese Journal of Computers 39(4)(2016) 780-794.

[9] M. Liu, G. Zeng, W. Yue, Y. Liu, Z. Qin, Online advertising click rate prediction model based on fusion structure, Chinese Journal of Computers 42(7)(2019) 1570-1587.

[10] Z. Zhang, Research of network click fraud and prevention strategy, [dissertation] Xinjiang: Xinjiang University, 2011.

[11] G.E. Hinton, S. Osindero, Y.W. Teh, A fast learning algorithm for deep belief nets, Neural Computation 18(7)(2006) 1527-1554.

[12] B. Hu, Y. Zhang, W. Chen, Characterizing search intent diversity into click models, in Proc. the 20th International Conference World Wide Web, 2011.

[13] J. Yang, Optimal strategy effect of online keywords auction and influence of budget, [dissertation] Shanghai: Shanghai Jiao Tong university, 2008.

[14] T. Khot, S. Natarajan, K. Kersting, Gradient-based boosting for statistical relational learning: the Markov logic network and missing data cases, Machine Learning 100(1)(2015) 75-100.

[15] H.B. McMahan, G. Holt, D. Sculley, Ad click prediction: A view from the trenches, in Proc. the19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2013.

[16] J. Yuan, An effective strategy to prevent clicking fraud, Journal of Computer Applications 29(7)(2009) 1790- 1793.

[17] S. Rendle, Factorization machines with lib FM, ACM Transactions on Intelligent Systems & Technology 3(3)(2012) 219-224.

[18] C.W. Coley, W.H. Green, K.F. Jensen, Machine learning in computer-aided synthesis planning, ACCOUNTS OF CHEMICAL RESEARCH 51(5)(2018), 1281-1289. DOI: 10.1021/acs.accounts.8b00087

[19] Z.L. Jiang, S. Gao, W. Dai, Research on CTR prediction for contextual advertising based on deep architecture model, Control Engineering and Applied Informatics 18(1)(2016) 11-19.

[20] P.G. Jing, Y.T. Su, L.Q. Nie, X. Bai, J. Liu, M. Wang, Low-rank multi-view embedding learning for micro-video popularity prediction, IEEE Transactions on Knowledge and Data Engineering 30(8)(2018) 1519-1532.

[21] J. Kang, R. Schwartz, J. Flickinger, S. Beriwal, Machine learning approaches for predicting radiation therapy outcomes: a clinician's perspective, International Journal of Radiation Oncology Biology Physics 93(5)(2015) 1127-1135.

[22] M.A. King, A.S. Abrahams, C.T. Ragsdale, Ensemble learning methods for pay-per-click campaign management, Expert Systems with Applications 42(10)(2015) 4818-4829.

[23] D. Lee, K. Hosanagar, H.S. Nair, Advertising content and consumer engagement on social media: evidence from facebook, Management Science 64(11)(2018) 5105-5131.

[24] S.C. Matz, J.I. Menges, D.J. Stillwell, H.A. Schwartz, Predicting individual-level income from Facebook profiles, PLOS ONE 14(3)(2019). DOI:10.1371/journal.pone.0214369.

[25] L. Miralles-Pechuan, D. Rosso, F. Jimenez, J.M. Garcia, A methodology based on deep learning for advert value calculation in CPM, CPC and CPA networks, Soft Computing 21(3SI)(2017) 651-665.

[26] L. Pang, S.A. Zhu, C.W. Ngo, Deep Multimodal learning for affective analysis and retrieval, IEEE Transactions on Multimedia, 17(11SI)(2015) 2008-2020.

[27] E.M. Schwartz, E.T. Bradlow, P.S. Fader, Customer acquisition via display advertising using multi-armed bandit experiments, Marketing Science 36(4)(2017) 500-522.

[28] R.F. Thompson, G. Valdes, C.D. Fuller, C.M. Carpenter, O. Morin, S. Aneja, W.D. Lindsay, H.J.W.L. Aets, B. Agrimson, C. Deville, S.A. Rosenthal, J.B. Yu, C.R. Thomas, Artificial intelligence in radiation oncology: A specialty-wide disruptive transformation? Radiother Oncol. 129(3)(2018) 421-426. DOI:10.1016/j.radonc.2018.05.030

[29] J. Wang, W. Zhang, S. Yuan, Display advertising with real-time bidding (RTB) and behavioural targeting, Foundations and Trends in Information Retrieval 11(4-5)(2017) 297-435.

[30] Q. Q. Wang, F.A. Liu, S.N. Xing, X.H. Zhao, T.L. Li, Research on CTR prediction based on deep learning, IEEE Access 7(2019) 12779-12789.