A Click-Through Rate Prediction Algorithm Based on Real-Time Advertising Data Logs

Chen Gong(✉)

School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu, China
(✉)Corresponding author: gongchen2020well@163.com

Abstract. Advertisement is one of the ways merchants attract consumers and enhance their influence. In recent years, under the big picture of smart city and smart life, the advertising business has shown a trend of specialization. It has always been a tough problem for advertisers of how to show potential consumers in real time the advertisements in a precise way. This passage reveals an experiment that alters the logistic regression function model and fits the data logs that users produce in order to forecast the possibility that the user clicks the real-time advertisement designed for him/her. The results of the experiment show that compared to traditional advertising, such altering improves the effectiveness of precise advertising, and has the edge of shorter operating time against other models based on deep learning. An aspect to improve is the comparatively lower accuracy of forecast. Such results indicate an alternative for certain small and medium enterprises to lower their advertising costs.

Keywords: click-through rate prediction, logistic regression, advertising data logs

1 Introduction

Real time advertising is a new category of advertisements in the background of smart city. Similar to the RTB real-time bidding system, it predicts users' preference for products by visiting massive operating data in applications and pushes relevant ads accordingly in real time. Different from traditional fixed advertisement page, this technique can make users more likely click into the relevant ads at a multiple-fold probability, so that it is more likely to discover products that attract users and even generate purchasing benefit to the selling company[1,2]. Secondly, for companies that need advertising, applying this technique enables precise advertising, instead of traditional fixed time and quantity massive advertising, lowering relevant costs. Thirdly, for advertisers, holding such technique ensures that every advertising company is duly willing in each respective time slots, increasing the advertisers' benefit. It will be an all-win situation[3].

1.1 Research on the techniques of click-through rate prediction

The major techniques with the click-through rate prediction model include clearing the data and differentiating training set and test set; choosing the proper learning model according to the characters of the data, e.g. logistic regression function model; learning the model parameters through the training set data; and testing the click-through rate prediction model using the test set data.

The technique of click-through rate prediction is developed to enhance the value of Internet advertising. A clicks ratio forecast model with less error could increase the margin from advertising and marketing.

The analysis of click-through rate prediction combines skills of management, information technology and modeling. It is designed for the data-intensive world nowadays. However, the forecasting technique brings along inevitable errors, which can sometimes affect sales of the products[4,6,8].

2 Constructing click-through rate prediction model for advertising data logs

2.1 Related mathematical functions and regularization

The mathematical form of the function is formula (1) below:

$$\text{sigmoid}(f(x)) = \frac{1}{1+e^{-f(x)}} \tag{1}$$

The corresponding curve is as figure1 below:

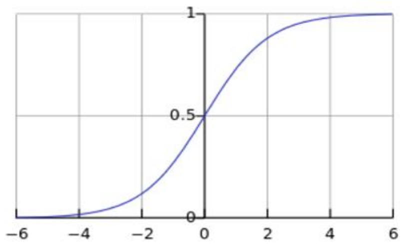

Fig 1. Function image

The function is very robust. It maps the input range of the function (-∞,∞) with the output range (0,1) and is probabilistic. Such character allows us to explain probabilistically.

The mathematical form of the model confirmed, problems are left with how to solve the parameters. A commonly used method in statistics is maximum likelihood estimation, during which a group of parameters should be found to maximize the likelihood (probability) of the data.

On the other hand, in the arena of machine learning, we more often encounter the concept of loss function. It measures the degree of model prediction error. Some common loss functions include 0-1 loss, log loss, hinge loss, etc., in which log loss is defined as formula (2) and (3):

$$p(y=1|\ x)=\frac{e^{w^T x+b}}{1+e^{w^T x+b}} \tag{2}$$

$$p(y=0|\ x)=\frac{1}{1+e^{w^T x+b}} \tag{3}$$

If an average of log loss of the whole dataset is obtained, we can infer that in a logical regression model, it is equal to maximize a likelihood function and to minimize a log loss function[5,9,10]. There exist multiple solutions to this optimization problem. Here, Gradient Descent is taken as an example. Gradient Descent, also called as Steepest Gradient Descent, is a method to provide iterative solution. It approaches the optimized value via selecting a direction adjusting parameter that makes the target function change the fastest in each step.

When there are too many parameters in the model, over-fitting issue often occurs. Hence, a way to control the complicatedness of the model will be needed. A typical method is to add regular terms into the optimization target, so that over-fitting is prevented by punishing the extra-large parameters.

Usually, let (p=1) or (p=2), to correspond with regularized L1 and L2, and the difference between the two can be seen in the many publications related to regularization. Regularizing L1 tends to make the parameter 0, thus a sparse solution is produced[7].

2.2 building discriminative model and clicks ratio model

Logical regression is a kind of discriminant model. It directly builds the model on conditional probability function P(y|x), without concerning the data distribution P(x,y) behind. Alternatively, the Gaussian Naive Bayes model is a kind of generative model. It builds a model on the joint distribution of data first, then calculates the posterior probability of the sample that belongs to all different categories with the Bayesian formula. This can be expressed as formula (4) below:

$$\mathrm{l}(w,b)=\sum_{i=1}^{m}\ln p\left(y_i \mid x_i;w,b\right) \tag{4}$$

Usually, it is hypothesized that P(x|y) is a Gaussian distribution, while P(y) is a polynomial distribution. The respective parameters can be obtained by maximum likelihood estimation. If we consider two category problems, we get the following with some simple deformations as formula (5) below:

$$\log\frac{P(y=1\mid x)}{P(y=0\mid x)}=\log\frac{P(x\mid y=1)}{P(x\mid y=0)}+\log\frac{P(y=1)}{P(y=0)}=-\frac{\left(x-\mu_1\right)^2}{2\sigma_1^2}+\frac{\left(x-\mu_0\right)^2}{2\sigma_0^2}+\theta_0 \tag{5}$$

We can see that the probability is the same as the form in logical regression. In this occasion, GNB and LR will learn the same model[11,12]. In fact, under a more common hypothesis (that P(x|y) is an index distribution), we can still arrive at similar conclusion.

To sum up, the mathematical model and solving of logical regression are both comparatively concise and easy to achieve. Through discretization of the character and other mapping, logical regression can also resolve non-linear problems. It is a powerful classifier. Therefore, in actual application, when we can obtain many lower tier characters, we may consider using logical regression to solve our problem.

What's more, in the course of running ads, the display position, time, attention on the products, gender of the user should also be concerned. In the stage of data analysis, the size of the dataset should be paid attention to, so as to avoid unnecessary loss.

There are three major methods of research and modeling for forecast analysis: traditional, data self-adaptive and model dependent. Academic research on traditional method of statistical inference and modeling starts from theory or model establishing. Statistical inference uses classical or the Bayesian method[7,8]. The traditional method, e.g. linear regression or logical regression estimates parameters for linear estimate variables. The model setup includes model fitting and model diagnosis. Before applying a traditional model for forecast, we test the model first.

To forecast, we may use the classical method or the Bayesian method, or we may totally avoid traditional statistics and machine learning models. We need to do effective research. Forecast analysis is based on a simple premise: the value of model is in the quality of its estimation.

To evaluate the quality of a clicks ratio model, there are various qualitative or quantitative, or online or offline methods. Yet, whatever evaluation method it is, the core is the same: to see the difference between clicked display and un-clicked display in this model. Of course, a quantitative index that can be calculated will be no better.

The model of logical regression is a non-linear model: the sigmoid function, or the logical regression function. But it is actually another linear regression model. Apart from the sigmoid mapping function relationship, all other steps and models are linear regressive. One may say that logical regression models are all supported by linear regression theoretically.

However, linear models cannot achieve the non-linear form of sigmoid, and sigmoid can handle the 0/1 classification problems easily. In addition, its derivative meaning is still the same as the maximum likelihood derivative of linear regression, the formula (6) and (7) are shown below

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m} y^{(i)} \log h_\theta\left(x^{(i)}\right) + \left(1-y^{(i)}\right)\log\left(1-h_\theta\left(x^{(i)}\right)\right)\right] \tag{6}$$

$$f(t) = \frac{e^t}{e^t+1} = \frac{1}{1+e^{-t}} \tag{7}$$

3 Results and testing

We refer to the quantitative unstableness of statistics. On one hand, we may use classical methods like confidence interval, point estimate with certain standard deviation, significance test and P value. On the other hand, we have the Bayesian statistics methods: posterior testing probability distribution, probability interval, estimate interval, the Bayesian factor and subjective (or perhaps diffusive) prior testing. Some of the evaluation index, such as the Akaike information criterion (AIC) or the Bayesian information criterion (BIC), is helpful in us determining the models, so as to balance between fitness and conciseness.

The essence of the method is the combination of training and testing. We divided the sample test data into training set and test set. The former is to build up the model while the latter is for evaluation. The dataset is divided into two to three parts, as shown in the following figure. The sample is divided into training set and test set randomly, especially when dealing with small datasets. Thus, we sometimes operate the statistics experiment through some random division and the average of performance index. There are some extension and changes in training and testing themes.

Finally, through improvements on the logical regression model and some parameters in the fitting process, the extract part of clicks ratio forecast by this model is shown as the table 1 as below:

Table1. Loss of this click-through prediction algorithm

Id	Position ID	Click time	Elapsed time	Click probability	Log loss
1164848	293	170000	0.51285	0.007586	0.090690
2127247	6161	170000	0.54658	0.006508	
2769125	7434	170000	0.52065	0.003381	
2513636	977	170000	0.58748	0.007815	
1144620	3688	170000	0.60015	0.010232	

As the table 1 shows, the click probability means likelihood of the next advertisement that customers may click through the data calculated by the algorithm. And log loss function represents the performance of the algorithm, generally speaking, prediction model is better when log loss is lower.

In order to verify the effectiveness of such model in real-life situation, I have operated some relevant tests, in which I picked randomly different participants, ran real time ads that they may be interested in through analyzing the ad data log they produced, and calculated the probability of them clicking into the ads. The extract part of results are shown in the table 2 as below.

Table 2. Tests on this model

ID of advertisement	Proportion of push messages	Average response time	Success rate
2127247	14.55%	84.082	85%
2769125	12.72%	71.005	80%
2513636	29.09%	56.258	95%
1144620	29.09%	46.725	96%

The above is a few representative advertisements of a broad category, they are pushed according to the click prediction rate obtained by the algorithm, and with different probability. I set the timeout time to 100 seconds, if the time goes out, the ad will be regarded as a failure to push. At last the result shows that about 90% of users responded to the push content within 100 seconds, so the prediction model is generally effective.

4 Conclusion

This method has a forecast accuracy of 91%, which is only several percentage points lower compared with other clicks ratio model of depth learning background. Yet, it operates more quickly, suitable for circumstances of real time pushing ads. This has also provided new inspirations for advertisers: to set different bidding price for different clicks ratio forecast probability, so that time slots can be utilized to the most, and that the bidding merchant can save advertising costs. Moreover, for many tourist cities, such technique can be used to push city icons to tourists more to their interest (e.g. scenic sites, food, culture, etc.). In this way, the destination city can attract more tourists. When data volume surges, how to store and manage data effectively, and when certain user data involves privacy issue, how to ensure the model safety will be suggested research areas.

Acknowledgements

I would like to acknowledge that this research is supported and funded by University of Electronic Science and Technology of China.

References

1. Haseon Park,Soojung Kim,Joonghwa Lee. Native advertising in mobile applications: Thinking styles and congruency as moderators[J]. Journal of Marketing Communications,2020,26(6).
2. The Advertising Research Foundation; ARF Announces Initiative To Advance Cross-Platform ROI Analysis Through Application Of Randomized Control Trial Measurement[J]. Computer Technology Journal,2020.
3. Wangbing Shen,Haixia Gu,Linden J. Ball,Yuan Yuan,Cheng Yu,Rong Shi,Taozhen Huang. The impact of advertising creativity, warning-based appeals and green dispositions on the attentional effectiveness of environmental advertisements[J]. Journal of Cleaner Production,2020,271.
4. Hao Li,Huichuan Duan,Yuanjie Zheng,Qianqian Wang,Yu Wang. A CTR prediction model based on user interest via attention mechanism[J]. Applied Intelligence: The International Journal of Research on Intelligent Systems for Real Life Complex Problems,2020,50(8).
5. Mohammad Malekahmadi,Sadegh Kazemi,Farideh Sharifipour,Farshad Ostadian,Atefeh Mahdian Rad,Mohammad Sadegh Mirdehghan.Effect of capsular tension ring implantation on predicted refractive error after cataract surgery in patients with pseudoexfoliation syndrome[J].International Journal of Ophthalmology,2020,13(04):587-590.
6. 김태석. Design of a Multi-array CNN Model for Improving CTR Prediction[J]. 한국콘텐츠학회논문지,2020,20(3).
7. Zhulin Tao,Xiang Wang,Xiangnan He,Xianglin Huang,Tat-Seng Chua. HoAFM: A High-order Attentive Factorization Machine for CTR Prediction[J]. Information Processing and Management,2019.
8. Jelena Gligorijevic,Djordje Gligorijevic,Ivan Stojkovic,Xiao Bai,Amit Goyal,Zoran Obradovic. Deeply supervised model for click-through rate prediction in sponsored search[J].

Data Mining and Knowledge Discovery,2019,33(5).

9. Jiang Zilong,Gao Shu,Li Mingjiang. An improved advertising CTR prediction approach based on the fuzzy deep neural network.[J]. PloS one,2018,13(5).

10. Lili Shan,Lei Lin,Chengjie Sun,Xiaolong Wang. Predicting ad click-through rates via feature-based fully coupled interaction tensor factorization[J]. Electronic Commerce Research and Applications,2016,16.

11. Zhe Gao,Qigang Gao. Ad-centric Model Discovery for Prediciting Ads's Click-through Rate[J]. Procedia Computer Science,2013,19.

12. Shimura T.,Harada J.. Absolute measurement of CTR scattering and comparison with theoretical predictions[J]. Shimura T.;Harada J.,1994,198(1-3).