The Study of User Characteristics Factors that Affect the CTR of ADs

Kangkai Zhao(✉)

Faculty of Science & Engineering, Curtin University, Kent St., Bentley, WA 6102, Australia
kangkai.zhao@student.curtin.edu.au

Abstract. With the continuous development of online shopping platforms, it is becoming more and more common for people to use e-commerce platforms for shopping; at the same time, e-commerce platforms pushing product ads to consumers have become one of the important ways to increase product sales. Therefore, the click-through rate (CTR) of product ads is getting more and more attention from decision-makers; how to effectively improve the CTR of ads has become a hot topic today. Previous studies on factors affecting CTR have basically focused on the characteristics of the ads, with less research on the different characteristics of consumers. This study used the platform ad click data collected by Taobao platform in 2017, and used regression analysis and machine learning to investigate the relationship between CTR and factors such as gender, age, consumption level, city level, recent behaviors etc.; The study found that consumers' gender, age, consumption level and recent behaviors have a significant impact on consumers' click-through decisions and CTR; this can be used as a basis to provide implications for future advertising strategies.

Keywords: Click-through Rate · Online Advertising · User Characteristic · User behavior

1 Introduction

1.1 Background

Since 1990, e-commerce has been developing on a large scale; and with the continuous development of the logistics industry and the popularity of the Internet, B2C and C2C e-commerce platforms have become one of the important channels for people to shop and consume in various countries. In the current e-commerce platform, it is not difficult to find that whether in the traditional web page, or mobile app page, the platform will not only provide a powerful product search function, but also push product advertising to the users actively; the e-commerce platform can also choose to push the product advertising in other webs, such as search engine listings and so on. At the same time, with the improvement of people's quality of life, consumers are gradually no longer limited to buying products they actively search for, more and more consumers are also going to click on the ads pushed by the platform. So, how to correctly push ads to different user groups in order to improve the click-through rate (CTR) of the ads has become a key topic for e-commerce platforms in recent years.

G. Ali et al. (Eds.): ISEMSS 2022, ASSEHR 687, pp. 3426–3435, 2022.
https://doi.org/10.2991/978-2-494069-31-2_402

1.2 Literature Review

The perspectives on the topic of ad click-through rate (CTR) decisions are multifaceted. The most mainstream research direction is to start from the active angle of advertising and study what types of ads are more popular with consumers. Azimi et al. found that visual features are efficacious in predicting click-through rates [1]. This suggests that different visual features of ads can have an impact on the CTR of ads. Sokolik et al. found that using red and other warm colors in web ads can generate dramatically higher CTR, and cool colors may cost an advertiser dearly in terms of CTR [2]. In addition, for the study of text in advertisements, in the article *Psychological Advertising: Exploring User Psychology for Click Prediction in Sponsored Search*, the authors stated that textual patterns catching user psychological desire can drive up the CTR significantly [3]. For specific styles of advertising, such as banner ads, there are also many corresponding studies. For example, Robinson et al. found that Size, message length, offer content etc. can all have a significant impact on ad CTR [4]. The second research direction is the study of keyword search ads. Keyword search ads are becoming more popular, and they comprise the largest Internet advertising revenue source, accounting for $8.8 billion or 41% of 2007 revenue [5]. It has been shown in previous studies that the order of ads placed in search engines affects the CTR of ads [6], and listing ads in decreasing order of their quality creates higher revenue [7]. Besides, there have been previous studies on CTR of ads for a specific product. Gómez-Carmona et al. had done research on wine ads on e-commerce platforms, the results show that whether or not the wine bottle is labeled in ads makes a difference in capturing consumers' attention and gaining sustained attention [8]; This may indicate that the best ads for different products may match different styles. Finally, there was also a passive study based on consumer characteristics, which was the most similar to this article. This direction of research is from the consumer's angle, studying the impact of different characteristics of users on the CTR of ads. Experiments in Gu et al.'s study shown that user habit, attitude and commercial intention have strong correlations with ad display session CTR [9]. Based on this perspective, more research could be done on whether different user characteristics (age, gender, consumption level, behavior, etc.) also have an impact on CTR of ads? The previous research in this direction has certain gaps.

1.3 Objective

This study will use regression analysis and machine learning to investigate the user characteristics and behavior factors that affect the CTR of ads. First, this article will organize the collected data in the method section; then, the hypothesis will be made according to the problem of study and the analysis will be carried out step by step; in the result section, the paper will describe the results of the data analysis, and the explanation of the results will be carried out in the discussion section later. At the same time, it is hoped that the results of this study will be useful for the ads push decision of e-commerce platforms.

Table 1. The data used and organized

Variable	Description	
User ID	Desensitized	
Product ID	Desensitized	
Whether to click	Yes (1)/No (0)	
Age group of users	0, 1, 2, 3, 4, 5, 6 (60–69 years old)	
User Gender	Male (1)/Female (2)	
The user's city level	1, 2, 3, 4	
Whether the user is a college student	Yes (1)/No (0)	
User's behavior in the last 22 days (For a particular product)	Whether to browse	Yes (1)/No (0)
	Whether to like	Yes (1)/No (0)
	Whether to add to cart	Yes (1)/No (0)
	Whether to buy	Yes (1)/No (0)
Price of the product in the ads	*	

2 Method

2.1 Data Collection and Organization

The dataset for this study was collected on the Ali Yun Tianchi platform [10], the dataset randomly sampled 1140000 users from the website of Taobao for 8 days of ad display/click logs (26 million records) in May 2017, as well as the characteristics of users who were pushed with ads, and their behaviors over the last 22 days. (China is the world's largest and most dynamic e-commerce market, ranking first in the world in terms of B2C sales and number of online consumers [11]. Alibaba, as the oldest e-commerce platform in China, occupies a very large share of the Chinese e-commerce market, and his B2C cum C2C platform ------ Taobao has been the first choice of most people for online shopping) User ID and product ID in the data had been desensitized. Also, some of the variables were categorized for the follow-up study. For example, in the variable of whether to click on the ad, 0 means no click, 1 means to click; in the variable of gender, 1 means male and 2 means female; all the variables that were categorized included: whether to click on the ads, gender, age, consumption level, city level, whether the user was a college student and user's behaviors. The data used and organized in this study are shown in Table 1.

2.2 Hypothesis

In order to facilitate the study and to make the objectives clear and structured, the following hypotheses were made for this study.

(1) There is a significant relation between different users' characteristics (user gender, age, city level, consumption level, whether they are college students) and users' decision to click on ads.

Table 2. The results of the covariance test

Variable	Tolerances	VIF
Price	1.000	1.000
Gender	0.985	1.015
Age	0.885	1.129
Consumption	0.942	1.061
Whether college	0.920	1.087
City level	0.983	1.018

(2) Different recent behaviors of users (within the last 22 days) affect the CTR of ads.

According to the hypothesis, the dataset was divided into two groups; one for hypothesis 1 was users' characteristics (user gender, age, city level, consumption level, whether they are college students) and whether or not to click on the ad; to facilitate machine learning and to classify consumers, the price of the product was also included. Another dataset (for hypothesis 2) included recent behaviors of users and CTR (per user for each product); the CTR is denoted as:

$$CTR_p = Numberoftimesclickedads/Numberoftimesreceivedads \tag{1}$$

2.3 Procedure

For hypothesis 1, before performing the logistic regression analysis, a covariance test was performed to ensure the stability of the resulting model and to exclude interference between variables (Table 2).

Dependent variable: Whether to click.

As shown in Table 4, the tolerances of all six independent variables are much greater than 0.1 and the VIFs are all much less than 10. This indicates that the correlations between the independent variables are not significant and are within an acceptable range.

The next step was to perform a logistic regression analysis of the data; 'whether to clicked'(0 means no click and 1 means click) was used as the dependent variable and the other six data items (user gender, age, city level, whether they are college students, consumption level, Price of the product) were used as covariates; all of them except the product price item were classified as categorized covariates and the benchmark value was the first indicator. The regression method is Forward: conditional. Regression equations were established.

$$\begin{aligned} \mathrm{WTC_i} = \mathrm{c} + \beta_0\mathrm{Price_i} + \beta_1\mathrm{Gender_i} + \beta_2\mathrm{Age_i} + \beta_3\mathrm{Consumption\ level_i} \\ + \beta_4\mathrm{Whether\ college_i} + \beta_5\mathrm{City\ level_i} + \mathrm{e_{it}} \end{aligned} \tag{2}$$

For a more general study of the data, a decision tree model by machine learning was built. The model used 'whether to click' as the dependent variable and the other six data as independent variables; the growth method used is CHAID; the proportion of learning samples was 70%, the proportion of testing samples was 30%; and the parent and child nodes in the minimum number of cases are 400/200, respectively.

Next was the study for hypothesis 2. Using the categorized behavioral data, linear regression equation was established with the dependent variable 'click-through rate of per user' and four behavioral variables.

$$\mathrm{CTR_p} = \mathrm{c} + \beta_0\mathrm{Browse_i} + \beta_1\mathrm{Like_i} + \beta_2\mathrm{Cart_i} + \beta_3\mathrm{Buy_i} + \mathrm{e_{it}} \quad (3)$$

3 Results

In the logistic regression with WTC as the dependent variable and the 6 user characteristics as covariates, the final results entered into the model are shown in Table 3.

In Table 3, Exp(B) is OR value, when the OR value is greater than 1, it means that this factor is more likely to make users click on the ads compared to the benchmark value. According to the p-value (Sig), gender, age, and consumption level have a significant impact on the user's decision to click on an ads(In the logistic regression, part of one categorical covariate is significant, which means the overall covariate significant). Women are more inclined to click on ads pushed by e-commerce platforms than men, and the ratio of clicks between men and women is about 1:1.071. In the sample of users aged 0–69, with the growth of age, the chance of users clicking on ads shows a curve of first decreasing and then increasing. Taking the user group of 0–9 years old as the benchmark, the data reach the lowest (0.895) in the age range of 20–29 years old, exceed the benchmark value of 1 (1.019) for the first time when the age of the user sample exceeds 50 years old, and reach the top (1.215) when the age of the user exceeds 60 years old (Fig. 1). Different consumption level of users can influence the decision to click on the ads. Among users in 3 different consumption levels (low, medium and high), the users' tendency to click on ads shows a decreasing trend as the consumption level rises. Using the users in the low consumption level as the benchmark value, the data are 0.989 and 0.890 in the medium and high consumption levels respectively (Fig. 2).

Meanwhile, the decision tree model shown the classification of users who clicked on the ads at different price segments. The ads pushed to the user are divided into five intervals according to the price of the product. In the range of prices less than or equal to 12.6 yuan, a total of 6.8% of users clicked, which was the highest proportion of users in the whole price segment; In addition, among users in the price (67.89,316] price range, the gender factor played an important role in the classification, the ratio of men to women was 1:2.023; and 67.29% of the female group in this price segment were in the middle to high consumption level. In other price segments, there was no clear classification.

In the linear regression analysis of CTR with user behaviors, three behaviors were significant (p-value < 0.05) except for Browse, as shown in Table 4.

Based on the data, building the model of CTR with like, add to cart and buy.

$$CTR_p = 0.381 - 0.051Like_i - 0.048\text{Add to cart}_i + 0.090Buy_i + e_{it} \quad (4)$$

Table 3. Variables in the equation

		B	Sig	Exp(B)
Steps 1a	Age (0)		0.000	
	Age (1)	−0.023	0.961	0.977
	Age (2)	−0.105	0.819	0.900
	Age (3)	−0.054	0.907	0.948
	Age (4)	−0.029	0.949	0.971
	Age (5)	0.004	0.994	1.004
	Age (6)	0.175	0.704	1.192
	Constants	−2.934	0.000	0.053
Steps 2b	Gender (2female)	0.073	0.000	1.076
	Age (0)		0.000	
	Age (1)	−0.029	0.949	0.971
	Age (2)	−0.114	0.804	0.892
	Age (3)	−0.062	0.893	0.940
	Age (4)	−0.032	0.944	0.968
	Age (5)	0.004	0.993	1.004
	Age (6)	0.177	0.701	1.194
	Constants	−2.977	0.000	0.051
Steps 3c	Gender (2female)	0.068	0.000	1.071
	Age (0)		0.000	1.00
	Age (1)	−0.029	0.949	0.971
	Age (2)	−0.111	0.809	0.895
	Age (3)	−0.051	0.912	0.951
	Age (4)	−0.018	0.968	0.982
	Age (5)	0.019	0.967	1.019
	Age (6)	0.195	0.672	1.215
	Consumption (1low)		0.000	1.00
	Consumption (2mid)	−0.011	0.510	0.989
	Consumption (3high)	−0.117	0.000	0.890
	Constants	−2.967	.000	0.051

4 Discussion

Several reports have shown that user habit, attitude and commercial intention have strong correlations with ad display session CTR, but few studies have demonstrated a correlation between other user characteristics, user behaviors and CTR. The present study was

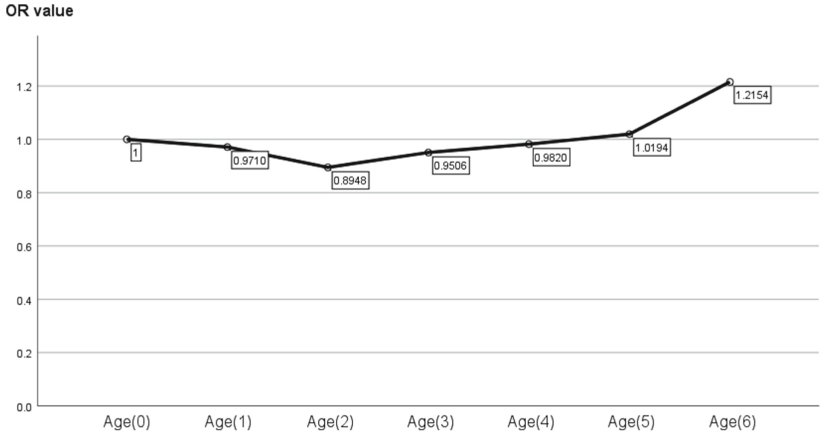

Fig. 1. OR value of Age group

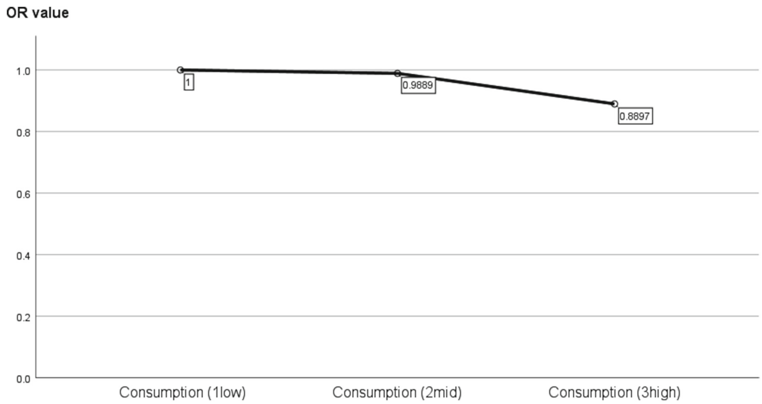

Fig. 2. OR value of Consumption level

Table 4. Regression of CTR with 4 user behaviors

	B	Sig	Tolerances	VIF
(Constants)	0.381	0.253		
Browse	0.307	0.357	0.999	1.001
Like	−0.051	0.000	0.996	1.004
Add to cart	−0.048	0.000	0.943	1.061
Buy	0.090	0.000	0.945	1.058

designed to determine the effect of some user characteristics, user behaviors on CTR of ads. One interesting finding is that the user's gender, age, and consumption level have a significant impact on the consumer's decision to click on an ad: women are more likely to click on ads than men, and consumers in the 50 and 60 age groups are most likely to click on an ad, while the age group 20 is the opposite. In addition, the higher the consumption level, the less tendency there is to click on an ad; these apparent differences may be related to different psychological factors between users when receiving an ad. And in the decision tree model with the same sample dataset, the $(0, 12.6]$ and $(67.89, 316]$ price segments deserve attention; the ads in the $(0, 12.6]$ segment have the highest CTR; and the ratio of male to female consumers who click on ads in the $(67.89, 316]$ price segment is about 1:2. Another important finding was that users' behaviors of like, add to cart, and buy over the last 22 days has a significant impact on their CTR of ads for a certain product. Among them, the behaviors of buy have the greatest impact, and it will prompt users to click on the ads; however, like and add to cart will reduce the CTR. A possible explanation for this might be that Users will stick to their original choice when they receive an ad for a familiar product. What is surprising is that browse behaviors does not have a significant impact on users' click-through rates; these results are likely to be related to the fact that more and more consumers prefer to 'wander' aimlessly through shopping platforms, so they have a very low impression of the products they have already browsed.

This finding has important implications for developing the CTR of ads on e-commerce platform. When pushing ads, decision makers can proactively push more ads to groups in higher age or lower consumption level; when pushing to women, they can increase the product proportion of $(67.89, 316]$ price segment. In addition, considering the overall cost of pushing ads, for products that users have already liked or have added to their shopping cart, the push of ads should be reduced.

However, there are some limitations should be considered in the future. The first is the timeliness of the data; the data of this study were collected in 2017, but considering the development speed of e-commerce platforms and the gradual upgrading of consumer perceptions, the results obtained from this study may have certain errors; second, this paper only investigated the influence of different factors on the CTR of ads; if the results of the study are to be applied to the pushing strategy of ads, there are many other factors should be considered; for example, it has been proven in studies that the behavior of buy can increase CTR, but the likelihood of consumers repeating the purchase is low; so the strategy of increasing ad push based on single behavior 'buy' is not appropriate. At the same time, this is an important issue for future research. Work is required to study the impact of more factors on ad push strategy, especially based on consumer psychological factors and the combined cost of ad push; further studies, which take these variables into account, will need to be undertaken.

5 Conclusion

This paper investigates the factors that influence users' click-through decisions on e-commerce platforms. It was found that consumers' gender, age, consumption level and recent behaviors have a significant impact on consumers' click-through decisions and CTR. Therefore, in the advertising strategy of e-commerce platform, the number of ads can be appropriately increased for high age and low consumption level consumer groups;

the number of ads can be increased for female consumers in specific price segments; meanwhile, ads that have the same product type should be avoided for users who have recently liked or added to cart.

References

1. J. Azimi, R. Zhang, Y. Zhou, V. Navalpakkam, J. Mao, X. Fern, Visual appearance of display ads and its effect on click through rate. Paper presented at the Proceedings of the 21st ACM international conference on Information and knowledge management, Maui, Hawaii, USA, 2012. https://doi.org/10.1145/2396761.2396826
2. K. Sokolik, R.G. Magee, J.D. Ivory, Red-Hot and Ice-Cold Web Ads: The Influence of Web Ads' Warm and Cool Colors on Click-Through Rates. Journal of Interactive Advertising, 14(1), 31-37, 2014. DOI:https://doi.org/10.1080/15252019.2014.907757
3. T. Wang, J. Bian, S. Liu, Y. Zhang, T.Y. Liu, Psychological advertising: exploring user psychology for click prediction in sponsored search. Paper presented at the Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, Chicago, Illinois, USA, 2013. DOI: https://doi.org/10.1145/2487575.2487699
4. H. Robinson, A. Wysocka, C. Hand, Internet advertising effectiveness. International Journal of Advertising, 26(4), 527-541, 2015. DOI: https://doi.org/10.1080/02650487.2007.11073031
5. C.Y. Yoo, The Effects of Persuasion Knowledge on Click-Through of Keyword Search Ads: Moderating Role of Search Task and Perceived Fairness. Journalism & Mass Communication Quarterly, 86(2), 401-418, 2019. DOI: https://doi.org/10.1177/107769900908600209
6. C.Y. Yoo, An Experimental Examination of Factors Affecting Click-Through of Keyword Search Ads. Journal of Current Issues & Research in Advertising, 33(1), 56-78, 2012. doi:https://doi.org/10.1080/10641734.2012.675559
7. A. Agarwal, T. Mukhopadhyay, The Impact of Competing Ads on Click Performance in Sponsored Search. Information Systems Research, 27(3), 538-557, 2016. DOI:https://doi.org/10.1287/isre.2016.0637
8. D. Gómez-Carmona, S. Cruces-Montes, P.P. Marín-Dueñas, C. Serrano-Domínguez, A. Paramio, A.Z. García, Do You See It Clearly? The Effect of Packaging and Label Format on Google Ads. Journal of Theoretical and Applied Electronic Commerce Research, 16(5), 1648-1666, 2021. doi:https://doi.org/10.3390/jtaer16050093
9. S. Gu, J. Yan, Z. Xiao, N. Liu, D. Shen, W. Fan, Y. Chen, What Are Driving Users to Click Ads? User Habit, Attitude, and Commercial Intention. In: ADKDD, 2010.
10. Ad Display/Click Data on Taobao.com. (2018, April 2). Tianchi-Dataset https://tianchi.aliyun.com/dataset/dataDetail?dataId=56
11. CIECC, World e-commerce report. Xinmin Weekly(16), 1, 2018.