ISSN: 2688-8653

Application of Data Analysis and Prediction in Advertising Delivery System

Xueying Hu[1,*], Zhuo Zhang[1], Jiahui Li[1], Ying Liu[1], Zejiong Zhou[2]

[1] School of Management Science and Engineering, Anhui University of Finance and Economics, Bengbu, Anhui, China

[2] School of Economics, Anhui University of Finance and Economics, Bengbu, Anhui, China

[*]Corresponding author email: 3088381411@qq.com

Abstract

The development of the Internet and the rise of e-commerce have brought great commercial value to more and more people, among which online marketing has become an important means of promoting brands, and more and more platforms have begun to build their own advertising delivery systems according to their own characteristics to undertake advertising and promotion to platform users. How to achieve accurate advertising and ensure the effectiveness of advertising marketing to protect the interests of advertisers while respecting the choice of platform users to achieve sustainable development is something we should care about. This paper takes Xiaohongshu as an example to collect its advertising data and conduct visual analysis.

Keywords

Data Analysis; Spark; Advertising Input System; Visualization.

1. Introduction

With the increasing coverage of personal computers and mobile phones, people can access a large amount of data in a short time and at a low cost, including the browsing of advertisements. This will generate a large amount of access data. However, how to use these data efficiently and then apply them to various business scenarios conducive to the company is still an unsolved problem [1]. The Financial Times report (2019) shows that the global revenue of online display advertising in 2019 has reached 144.7 billion dollars. In China, according to statistics, the market size of China's Internet advertising industry will reach 543.93 billion yuan in 2020, with a year-on-year growth of 12.6%, and 655.01 billion yuan in 2021[2]. It can be seen from this that under the premise of both the Internet scale and the Internet advertising industry market scale expanding, if the huge amount of data generated by users' browsing can be analyzed and processed in detail and used reasonably, the company's business development will be greatly promoted. In this regard, this paper intends to use a large amount of data from various companies for visual processing and export to form an advertising data analysis system, which makes it easier to carry out subsequent refined advertising operations after the data is collated and analyzed, and has important practical and commercial value.

The effectiveness of advertising is directly related to the accuracy of audience positioning. The more you know about the audience's individual information, consumption information and demand information, the closer your advertising activities will be to their needs and the more accurate your advertising will be [3]. In the context of the Internet, the boundary between the information disseminator and the audience has changed from a relatively single media tool to a diversified media equipment, becoming more opaque. At the same time, the communicator's information collection of the audience has become "transparent", and the third-party platform between the two will circulate their information. However, such data advertisers are difficult

ISSN: 2688-8653

to access directly, so they need to use third-party advertising platforms to deliver. The advertising delivery system is mainly used by advertising service intermediate operators and ordinary medium and large-scale websites. The advertising service intermediate operators have huge advertising customer resources. On the one hand, they need a system to facilitate advertising and management. On the other hand, they also act like a wholesaler, distributing advertisements to the next layer of websites, and so on. Therefore, these websites also need to be able to track and count the specific situations such as the reading amount of advertising.

Ordinary medium and large-scale websites often have certain strength, and can receive advertisements from large advertising service intermediate operators and continue to put them. With the increase of business volume, they need the advertising system to manage. If more in-depth data analysis is carried out, advertising can be carried out more accurately, and a lot of costs can be saved. However, the hits or views of many websites are mixed with "moisture", which has a great impact on the accuracy of data analysis. Therefore, accurate data and reasonable data analysis methods are very important. The most common method for detecting outliers is data visualization, such as Box plot, Histogram, and Scatter plot. Researchers can intuitively see which data exceptions are so that they can process the abnormal data to ensure the accuracy of data analysis. Common methods for handling abnormal data include deletion, exchange of combined values, estimation and separate processing. Typical data analysis algorithms include K-Means for clustering, SVM for statistical learning and Naive Bayes for classification. The main tools used are Hadoop's Mahout[4]. For this, different processing methods will be selected according to different situations.

2. Demand Analysis

As an advertiser, its main purpose is to promote its own products, improve product awareness, and thus improve product sales. As the advertiser, the goal is to get the most accurate delivery with the least cost, so as to get the maximum benefit. Advertisers can obtain users' preferences through the platform to achieve more accurate advertising.

User portrait is a concept put forward by Alan Cooper, the father of interaction design. It is defined as "user information labeling", and can also be understood as a business panorama of a user after an enterprise collects and analyzes the data of main consumer information. This information mainly includes social attributes, living habits, consumer behavior, etc., which is sufficient to provide enterprises with sufficient information base and help enterprises quickly find more extensive feedback information such as accurate user groups and user needs[5]. As the platform side, our main research goal is to analyze the current platform side's operating mechanism, analyze the different needs of different users according to user feedback and relevant data, and then consider the changes in user preferences and needs, closely follow the trend of environmental popularity, build user portraits, build a universal model, and achieve accurate delivery. The platform side also needs to select a suitable time period, geographical location, audience and other single attributes or a combination of multiple attributes to place advertisements.

From the perspective of users, our main research goal is to analyze users' preferences and needs based on their browsing history and search keywords, so as to accurately push advertisements for the products they need to avoid too many advertisements affecting their use experience. It is undeniable that this way provides convenience for users, and brings advertisers closer to customers, almost reaching the "face-to-face" service level. However, in most cases, users will feel that they are constantly being monitored by an invisible pair of eyes, and have no privacy [6]. What we need to do is to make clear to users the scope of information acquisition, and give them the right to choose to collect data within a legitimate and reasonable range to push advertisements.

ISSN: 2688-8653

In general, the main research objectives of this project are: to analyze the current advertising mode, collect accurate data, screen abnormal data, accurately analyze data, establish a universal model, and accurately launch advertising, so as to achieve accurate audience, accurate cost, and accurate effect, and achieve a win-win situation for advertisers, platforms, and users.

3. Data Visualization

We randomly sampled 150000 users' advertisement display/click logs within 7 days from the Little Red Book website to form the original sample skeleton.

Table 1. Overview of Data and Fields

Data name	Description	Attribute
Raw-sample	Original sample skeleton	User ID, advertisement ID, time, click or not
Ad-feature	Basic information of advertisement	Advertising ID, category ID
User-profile	User's basic information	User ID, age group, gender, city
Raw-behavior-log	User's behavior log	User ID, time, commodity category ID

Note: One advertising ID corresponds to one commodity. The field time is the time point when the user clicks the advertisement, in hours.

Table 2. List of Users and Advertising Information

User_id	Cate	Brand	Gende	Age_level	Class	Noclk
ud_1	Fashion	Romon	Female	25-34	Jiangsu	1
ud_2	Make up	Chanel	Female	25-34	Shanghai	0
ud_3	Fashion	Home of Hailan	Male	35-44	Anhui	1

Visualize the data after data cleaning such as missing value processing, missing value filling and data merging.

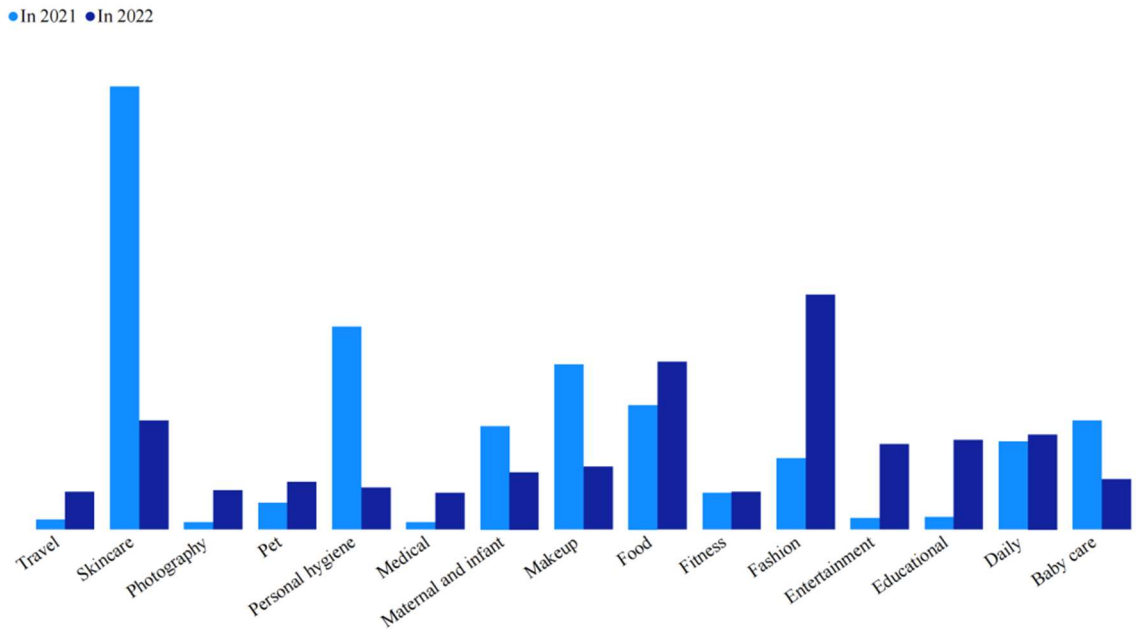

Figure 1. Advertising volume and category

ISSN: 2688-8653

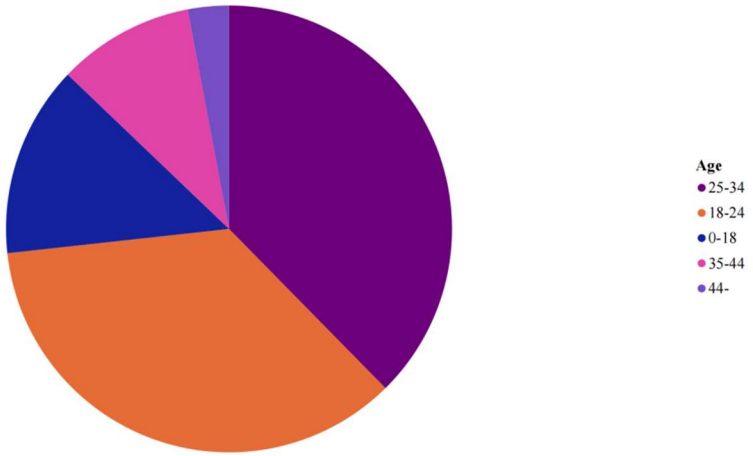

Figure 2. Advertising volume and age

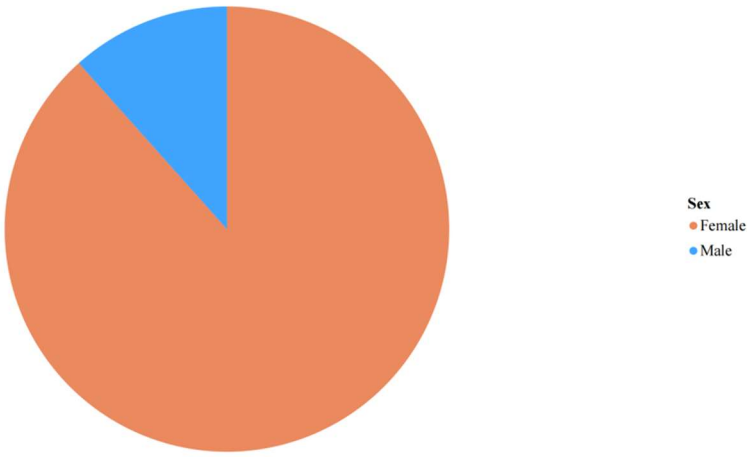

Figure 3. Advertising and gender

As an e-commerce platform perpendicular to women's content, Little Red Book is a further development and promotion of the e-commerce industry. It is also a new platform for brands, with emphasis on transformation and product promotion. It can be seen from the above chart that the advertising of Little Red Book mainly focuses on skin care, makeup and washing, and the mother and baby, cute pet and food also account for a certain proportion. Among them, the advertising volume of skin care and washing has increased significantly compared with 2021. According to the age of users, the main target audience is female users aged 18-34, accounting for about 70% of the total. It can be seen that they have a clear intention to pursue quality of life, are keen on rich quality of life, and have a strong intention to pursue consumption and taste new ideas. Such users have strong comprehensive consumption ability and high purchasing power. They are the leading consumers in the future and have huge market potential.

ISSN: 2688-8653

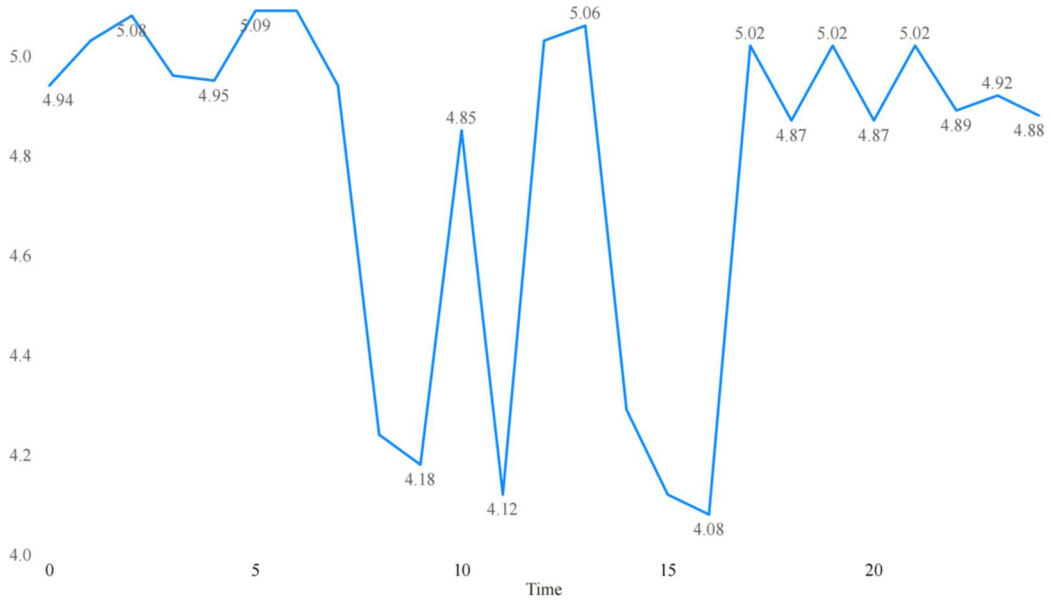

Figure 4. User click and time

It can be seen from the above chart that a large number of user clicks are concentrated in the evening and lunch break, and the amount of advertising at this time also increases.

4. Suggestions and Improvements

4.1. Platform

4.1.1. Add Release Management Module and Account Management Module

After advertising, the platform can see the advertising effect of all advertisements in the system through the advertising management module, including the characteristic information of the audience, so as to know whether the advertising matching is correct. You can also modify the created advertising information in this module. The account management module can help the platform view the account list information of all advertisers and merchants in the current advertising system, including email, company, account type, status and operation. At the same time, users should also add the option to turn off advertising. For example, the business model of China's integrated video platform is mostly a combination of commercial advertising and user payment. The platform should also respect users' choices to achieve sustainable development when cooperating with advertisers to deliver accurate advertising.

4.1.2. Control Platform Flow Quality

Comparative analysis: Compare the advertising traffic and non advertising traffic. Through this method, you can know whether it is the channel quality or the website itself.

Regularity of website traffic curve: understand users' habits of logging in to the website, and control the time period of advertising.

Time period change of website traffic: discover the rules and abnormal points of traffic, and then find the underlying causes, timely find problems, and adjust the delivery.

Establish composite indicators: decompose user behavior into data for evaluation.

A/B test: make two plans for the same optimization target (conversion rate), let some users use Plan A and some users use Plan B, and compare the conversion rate, clicks and other indicators

ISSN: 2688-8653

generated under different plans, so as to judge the advantages and disadvantages of different plans and make decisions to optimize the target.

4.1.3. Data Association of Each Platform

Collect users' browsing records in different APPs within a reasonable range to achieve accurate data. Use the Apriori algorithm to correlate the data of various platforms, maximize feedback on the content that users are interested in, and achieve accurate delivery. Association analysis is used to discover the connection and degree of connection between users' attention to a certain product or products on different platforms. For example, if a user browses the information or comments on a certain product on a certain platform, she is most likely to place an order on the shopping platform.

4.2. Users

4.2.1. Analysis of "Fragmentation"

The whole process is divided into user behavior path analysis and user profile analysis. The user's path analysis generates the degree of convergence by recording the user's statements on different platforms. The user profile analysis is based on detailed records of typical user groups. For example, click rate analysis of people of all ages, click rate analysis of people of all consumption levels, click rate analysis of people of all shopping depths, click rate analysis of people of all occupations (whether college students or not), click rate analysis of people of all city levels, etc. To achieve personalized marketing for different groups, relevant data analysis methods are used to analyze the interests of consumer groups, media usage, etc., so as to achieve the purpose of advertising and improve the rate and quality of advertising.

4.2.2. Physical Situation Tracking and "User Communication Network"

Physical situations refer to those environmental factors that exist in the form of physical objects and can arouse people's feelings. The consumption process of consumers is not only carried out in certain situations, but also affected by situational factors to varying degrees [7]. The platform can obtain the specific information of users when allowed by laws and regulations, and use positioning, distance sensors, etc. to learn the location information of users, and push appropriate advertisements for users according to the physical situation obtained and their consumption habits.

Advertising communication sometimes depends not only on the platform, but also on the communication between users. Use the "user communication network" to place different advertisements according to different circles of different users, and then spread the advertisements through the user's associated network, so as to achieve the minimum cost and obtain the maximum communication range. Moreover, the "user communication network" takes into account the social attributes of people, combines users' interests with their social relationships, builds a social network with users as nodes and replies and forwarding relationships as edges, divides the community structure and analyzes the group characteristics, calculates the point-to-point centrality to find the core nodes of the community, so as to determine the location and target group of advertising [8].

4.3. Advertisers

4.3.1. Advertising Development Model

Taking into account the development risks, the development model at each stage, the market direction and how to respond to changes in users' preferences, predict its future development model and record the development suggestions at each stage. Thus, the total communication efficiency of each advertising is the largest and the advertising budget cost is the lowest. After placing the advertisement, the advertiser can see the effect of the advertisement, and

ISSN: 2688-8653

understand the display volume, browsing volume and attention volume of a single advertisement from multiple perspectives.

4.3.2. Media Display Style Analysis

The form of new media advertising integrates all categories of traditional media, and can achieve the distribution of multiple forms of advertising within a resource category project. To a large extent, the delivery efficiency and accuracy are higher. The same channel often has a variety of presentation styles. Different presentation styles have a great impact on the advertising effect. This requires us to analyze and sort it out.

4.3.3. Advertising Model

Use big data analysis and visualization technology to analyze and screen the collected data, further analyze the "dehydrated" data, and use clustering K-Means, SVM and other algorithms to fit a set of advertising model. Different from the past, this model is not static, but will take into account changes in users' preferences and social mainstream trends, use KNN, Naive Bayes, Apriori and other algorithms to predict data, and record development suggestions at each stage, so as to maximize the overall communication efficiency of each advertising and minimize the advertising budget.

Acknowledgments

This work is supported by 2021 Anhui University of Finance and Economics Provincial Undergraduate Innovation and Entrepreneurship Training Program (No.: S202110378097).

References

[1] Lin Sui, Zhao Fei. Research on the application of Spark based linear model in advertising delivery system [J]. Journal of Guangdong University of Technology, 2016, 33 (5): 28-33.

[2] Dai Wenqiang, Chu Weijia, Zhong Jing. Optimization of online display advertising strategy of guaranteed volume contract under CPC mode [J]. China Management Science, 2022 (10): 1-11.

[3] Huang Mengfan. Research on advertising strategy based on big data [D]. Hunan University, 2018.

[4] Zhong Fanghong, Lin Yizhen, Shu Meng. Big data analysis based on comprehensive detection platform of cigarette packaging machine [J]. China Management Informatization, 2018, 21 (21): 80-82.

[5] Liu Zijing, Yu Runtong. Design of accurate advertising delivery system based on user portrait [J]. Jiangsu Science and Technology Information, 2022, 39 (10): 69-72.

[6] Zhong Rushi. Problems and Countermeasures of Information Flow Advertising [J] Cooperative Economy and Technology, 2022 (17): 72-73.

[7] Wu Congzhi. Analysis of the impact of situational factors on consumer psychology [J]. Journal of Mudanjiang University, 2008, 17 (12): 64-65.

[8] Wang Yan. Research on Precision Advertising Strategy Based on Social Network [D]. Central China Normal University, 2015.