WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 1, Number 1, Year 2023
DOI: https://doi.org/10.62051/ijcsit.v1n1.03
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Based on intelligent advertising recommendation and abnormal advertising monitoring system in the field of machine learning

Bo Liu *

Software Engineering, Zhejiang University, Hangzhou, China
*Corresponding Author

ABSTRACT

With the rapid development of the Internet, the scale of the online advertising market has expanded rapidly, and display advertising has become the most popular means of publicity. Accurate advertising recommendation is the guarantee of Internet platform revenue, and accurate advertising click rate prediction is the premise of accurate recommendation and abnormal advertising detection. Therefore, such monitoring and recommendation can be achieved through machine learning combined with artificial intelligence, and the application of intelligent AD recommendation systems and abnormal AD monitoring in the field of machine learning represents a complex integration of technologies to improve the precision and effectiveness of digital marketing strategies. Intelligent AD recommendation systems utilize advanced machine learning algorithms to analyze user behavior and preferences to deliver tailored AD content. These systems leverage vast amounts of user data, including browsing history, purchase history, and engagement metrics, to predict and present the most relevant ads. This paper analyzes the data mining in machine learning algorithms and the real-time online recommendation algorithm of Gaussian process, and analyzes the abnormal advertising monitoring system for maintaining the integrity and efficiency of advertising campaigns. By using machine learning technology for pattern recognition and anomaly detection, various measures and indicators of advertising campaigns can be monitored in a vigilant manner.

KEYWORDS

Intelligent advertising recommendation; Click rate prediction; Abnormal advertising monitoring; Machine learning.

1. INTRODUCTION

At present, Internet advertising is mainly divided into two categories from the exposure mode of search advertising and display advertising. Among them, the display of search ads is mainly based on the user's query words on the browser, combined with certain user click history records, and its advertising information is displayed to the user along with the search results. These ads contain explicit information about the user's request, so the results displayed by the browser are generally better suited to the user's needs. Display advertising is an extensive type of delivery in the absence of a clear request from the user, that is, the user does not provide keywords. This type of advertising has a larger market share than search advertising, but the economic conversion rate is very low. The traditional billing advertising recommendation form is that after the user realizes the payment, the consumer clicks on the advertisement, and then there is the possibility of the next browsing, purchasing and other behaviors, which can bring benefits to the advertising platform and the advertiser [1]. In the new intelligent advertising recommendation and advertising anomaly detection, the logistic regression algorithm of machine learning has a large number of applications in the industrialization of advertising recommendation, which regards each feature as an independent attribute, and assigns different weights to each attribute. In this paper, based on the machine learning model algorithm, a real-time recommendation algorithm based on Gaussian process is proposed for intelligent advertising online recommendation. In order to solve the universality problem caused by the assumption of linear relationship between product features and user feedback in traditional Lin UCB series online recommendation algorithms, a Gaussian process is proposed to describe the relationship between the two. Gauss processes are model-free algorithms that transform concrete functions into distributed representations.

It can fit a variety of complex functional relationships, so as to solve the universality of specific functional assumptions in various environments[2]. Aiming at advertising anomaly detection, a random forest-based advertising traffic detection method is proposed. By comparing the advertising traffic with the natural traffic, we get the characteristics of the natural traffic, construct the advertising traffic classifier, and get a good classification effect. The category object is a single traffic. Aggregate analysis of traffic is not required. Therefore, the storage capability of the device is reduced.

2. RELATED WORK

Internet advertising recommendation technology has always been a research problem of great commercial value. In practical applications, different forms of recommendation are often used for online and offline scenarios, and discussions are conducted regardless of the two different application scenarios[3].

1). Online recommendation system: It is suitable for the recommendation system platform with fast product update speed. The online recommendation system will not conduct large-scale modeling and training of historical data, but try to recommend users and update the recommendation strategy according to the immediate feedback of users to improve the click rate of users. The key to affect the accuracy of online recommendation system is the effective recommendation strategy.

2). Abnormal advertising detection: Through comparative analysis of advertising traffic and natural traffic, random forest has the advantages of processing high-dimensional data without feature selection and balancing errors. In this chapter, random forest algorithm is selected as the classification algorithm of advertising traffic. By constructing a random forest model and using the error rate outside the bag for internal evaluation and parameter adjustment, a random forest-based advertising traffic classifier is constructed

2.1. Online advertising intelligent recommendation technology

Reinforcement Learning (RL), also known as evaluative learning, is a paradigm and methodology of machine learning. RL is used to describe and solve the problem that an agent uses learning strategies to maximize returns or achieve specific objectives in the process of interacting with the environment. In recent years, a large number of scholars have added reinforcement learning to online recommendation systems[4-6]. Multi-Armed Bandit (MAB) is a kind of simple but very practical and widely used algorithm in reinforcement learning. In the multi-alarm slot machine problem, it is assumed that there are a fixed number of candidate arms, and the arm selected in each recommendation round will generate a random reward accordingly. The goal of the problem is to find the arm with the highest reward expectation and maximize the cumulative reward throughout the process. The biggest problem of MAB is the famous problem of exploration & exploit (E&E) in the recommendation system. Exploration refers to exploring new unknown information to fit a better recommendation model, and utilization refers to using known product information to choose an optimal choice under current conditions. In order to solve the E&E problem, Many algorithms based on Bandit have appeared at home and abroad.

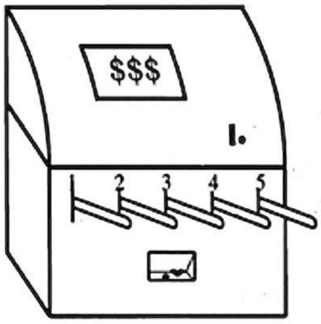

Figure 1. Multi-Armed Bandit (MAB) model

Context firee-Bandit algorithm mainly includes E-greedy algorithm, Thompson sampling and UCB algorithm. Watkins et al. proposed the E-GREEDY algorithm, which takes the probability of e to explore and the probability of 1-e to exploit, where E is the number between (0,1). Based on the E-greedy algorithm, many scholars have improved it. Even-DarE et al. proposed an e-priority algorithm, which randomly explored En times in the recommendation round, using the original best recommendation (1-n) times. Later, Kawale J et al. combined Bayesian probability decomposition method with Thompson sampling, and achieved good results [2. Chapelle and Li et al. applied Thompson algorithm to the fields of news and advertising recommendation, 127 and both achieved good results, proving the practical applicability of Thompson sampling in recommendation algorithms. Compared with the above algorithm, UCB[7] algorithm does not use random selection strategy, but estimates the return of existing projects, calculates the confidence interval of each project, and recommends the project with the largest confidence interval in each recommendation process.

2.2. Abnormal advertising detection technology

Anomaly detection refers to the identification of rare observation data that has an extreme value completely different from other data points. Such data are called outliers and need to be distinguished from the identification of the subject. The causes of these anomalies are often[8]: deviations in the data, errors in the data collection process or the emergence of new rare circumstances. Managing these outliers is often very difficult because it is impossible to understand whether the problem is related to faulty data collection or something else. Depending on the needs of our goals, decide whether to eliminate or maintain this outlier. If the outlier occurs due to the occurrence of a new event, deleting the outlier means losing information. Because in this case, the outlier contains important new information due to its rarity.

$$\mathrm{H}(x) = \arg\max_{Y} \sum_{i=1}^{k} I(h_i(x) = Y) \tag{1}$$

Random forest is a machine learning algorithm proposed by Leo Breiman (2001), which can be used for both classification problems and regression problems. Because of its high flexibility, random forest algorithm has broad application prospects, such as the prediction of diseases in medical treatment, the prediction of sales strategy in business and other analysis and prediction scenarios[9]. Compared with machine learning methods such as neural network (NN), support vector machine (SVM), decision tree (DT) and Adaboost, random forest algorithm can improve the prediction accuracy without significantly increasing the amount of computation, and is insensitive to multivariate collinearity, modeling missing data and unbalanced data is relatively robust, and the independent variables can be as many as several thousand. Known as one of the best algorithms at present, it is increasingly widely used.

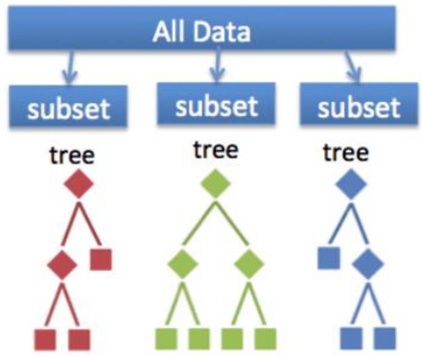

Figure 2. Random forest classification detection model

In this chapter, the random forest is constructed by CART tree algorithm, and the data set is divided by binary segmentation in CART tree. For numerical data, the sample set can be divided into two parts by selecting an optimal segmentation point. For nominal data with only two values, it is also possible to divide the sample set into two parts by treating each value as a branch. However, for nominal data whose number of values is greater than or equal to 3, it cannot be directly divided, so it is necessary to transform the nominal data.

3. METHODOLOGY

UCB algorithm is a classic gambling machine algorithm, through continuous exploration and utilization to find the optimal arm to maximize the reward. However, the problem of UCB algorithm is that the set is fixed, and the content of the UCB algorithm is dynamic in many application scenarios, so the algorithm cannot effectively recommend in a dynamic environment. In the context scenario, the algorithm makes a decision according to the characteristics of the item to be recommended, and adjusts the overall decision-making strategy based on the user's click feedback to maximize the user's click effect[10]. For newly generated items, the algorithm can process their features in real time, so that they can be added to the recommendation decision sequence.

3.1. LinUCB algorithm assumption

The LinUCB algorithm assumes that each item x corresponds to a D-dimensional feature vector xR, and assumes that there is a linear relationship between the item's expected reward r and its feature vector, which can be represented by the coefficient vector o*eR corresponding to item x. This paper assumes that the eigenvector of item x at time: is X and the reward at time is, then the expected reward satisfies the formula：

$$\mathrm{E}[r_{x,t} \mid x_t] = \theta_x^{*T} x_t \tag{2}$$

Suppose that at time t, 4 is a d x m item feature matrix, and each column corresponds to an item feature vector; b, is an M-dimensional vector that corresponds to the feedback received by each project. LiUCB algorithm uses ridge regression to solve this equation, and the estimated coefficient vector is as follows:

$$\theta_x = (A_t A_t^T + I_d)^{-1} A_t b_t \tag{3}$$

Id is a d*d identity matrix, and combined with the probability theory formula to achieve the accurate algorithm of online advertising intelligent recommendation.

$$|\theta_x^T x_t \text{-E}[r_{x,t} \mid x_t]| \leq \alpha \sqrt{x_t^T (A_t A_t^T + I_d)^{-1} x_t} \tag{4}$$

Among them, the first item uses the linear model to estimate the expected reward. The coefficient vector θx is calculated by using the interactive data, and then the expected feedback θTx,x under the current knowledge is calculated by combining the project characteristics. This step can be considered as the utilization process; the second item can be considered as the exploration process; the upper confidence interval is used to control the exploration degree of the algorithm. As the interaction with the user increases and the model learns more and more, the predictive power of the model will become stronger and stronger, and this item will become smaller and smaller[9-10]. Although LinUCB builds learning models based on projects, it models θx for each project x and modifies parameters θTx,x through continuous interaction with user feedback. However, the characteristics that the algorithm observes each time are actually user characteristics that are observed from a project perspective. After that, for each project to be recommended, the expected reward that the target user can get and the upper bound of the confidence interval are calculated, and the appropriate project is selected and recommended to the user considering both aspects.

3.2. Existing problem

The simple return situation of slot machines is determined internally by the slot machines themselves, while in the field of advertising recommendation, the return of a choice is determined by User and Item together. If we can use feature to characterize the pair CP of User and Item, before each item is selected, The expected return and confidence interval of each arm (item) can be estimated through feature, and the selected returns can be generalized to different items through feature.

UCB algorithm inserted the feature wings, which is the biggest feature of LinUCB.

LinUCB algorithm makes an assumption: After an Item is selected and pushed to a User, its return is linearly related to the relevant Feature, where the "relevant feature" is the context, and it is also the part with the largest space in the actual project.

Therefore, the test process becomes: use the characteristics of User and Item to estimate the return and its confidence interval, select the item recommendation with the largest upper bound of the confidence interval, and update the parameters of the linear relationship after observing the return, so as to achieve the purpose of experimental learning. The basic LinUCB algorithm is described as follows:

Inputs: $\alpha \in \mathbb{R}_+$
for $t = 1, 2, 3, \ldots, T$ **do**
  Observe features of all arms $a \in \mathcal{A}_t$: $\mathbf{x}_{t,a} \in \mathbb{R}^d$
  for all $a \in \mathcal{A}_t$ **do**
    if a is new **then**
      $\mathbf{A}_a \leftarrow \mathbf{I}_d$ (d-dimensional identity matrix)
      $\mathbf{b}_a \leftarrow \mathbf{0}_{d\times 1}$ (d-dimensional zero vector)
    end if
    $\hat{\boldsymbol{\theta}}_a \leftarrow \mathbf{A}_a^{-1}\mathbf{b}_a$
    $p_{t,a} \leftarrow \hat{\boldsymbol{\theta}}_a^\top \mathbf{x}_{t,a} + \alpha\sqrt{\mathbf{x}_{t,a}^\top \mathbf{A}_a^{-1} \mathbf{x}_{t,a}}$
  end for
  Choose arm $a_t = \arg\max_{a\in\mathcal{A}_t} p_{t,a}$ with ties broken arbitrarily, and observe a real-valued payoff r_t
  $\mathbf{A}_{a_t} \leftarrow \mathbf{A}_{a_t} + \mathbf{x}_{t,a_t}\mathbf{x}_{t,a_t}^\top$
  $\mathbf{b}_{a_t} \leftarrow \mathbf{b}_{a_t} + r_t\mathbf{x}_{t,a_t}$
end for

Figure 3. LinUCB Basic algorithm parameters

Among them, inUCB =L algorithm has a very important step to realize online intelligent recommendation of advertising, that is, to build features for User and Item, that is, to describe context, respectively

(1) Demography: gender characteristics (2 categories), age characteristics (discrete into 10 intervals)

(2) Geographical information: cities all over the world, states in the United States

(3) Behavior category: 1000 categories that represent the user's historical behavior

And finally you can normalize the original eigenvectors to unit vectors. It is also necessary to reduce the dimension of the original features, and the model must be able to describe some nonlinear relations. Logistic Regression is used to fit users.

There are many ways to do Explore, bandit algorithm is one of the schools. Several bandit algorithms have also been introduced before, which are basically the method of estimating confidence intervals, and then making recommendations according to the upper bound of the confidence interval, represented by UCB and LinUCB.

It is the combination of bandit and collaborative filtering algorithm, the difference between the two articles is that the latter only considers the User clustering (that is, only user-based collaborative filtering), while the former adopts co-clustering (co-clustering). It can be understood as two cooperative ways of item-based and user-based at the same time, and the latter is a special case of the former

4. CONCLUSION

With the rapid development of the Internet, the research and application of advertising recommendation algorithm has become a hot spot. In practical application, the characteristics of Internet advertising are high dimension, sparse distribution, fast update and so on. According to the requirements of product update rate and real-time performance, advertising recommendation technology is further divided into offline recommendation algorithm based on building training model and online recommendation algorithm based on decision. How to make good use of the characteristics of users and products to provide users with the most reasonable advertising recommendation has become a difficult problem. The research work of this paper is mainly based on the prediction of advertisement click rate and abnormal advertisement detection[11]. By analyzing the characteristics of linUCB algorithm, the data set is pre-processed. This paper proposes an online advertising click-through rate prediction algorithm based on attention mechanism and deep neural network, and an online advertising recommendation algorithm based on Gauss process. All black box functions can be fitted by analyzing Gaussian process, which makes the online decision algorithm not limited to the specific model representation, and enhances the universality of the online interactive algorithm. At the decision-making stage, two specific recommendation strategies are proposed based on the ideas of upper confidence interval and expectation maximization improvement, and the E&E problem is expressed.

The online advertising intelligent recommendation algorithm developed by the combination of artificial intelligence[12] (AI) and deep learning will further enhance personalization and precision in the future, so that advertisements are more in line with users' interests and needs. As the technology evolves, these algorithms are able to analyze huge amounts of data in real time, including user behavior, preferences, and interaction history, to predict what ads are most likely to piqued user interest. In the future, these algorithms will be more transparent and interpretable, while also paying more attention to protecting user privacy. In addition, they will continue to optimize themselves through enhanced learning capabilities to adapt to changing markets and user behavior, providing more accurate and efficient advertising solutions.

ACKNOWLEDGEMENT

This study is deeply grateful to authors such as Wan Weixiang, Jingyu Xu for their contributions to the field of intelligent healthcare decision support systems. Their article, "Development and Evaluation of Intelligent Medical Decision Support Systems," appears in the Academic Journal of

Science and Technology "Technology" provides us with extremely valuable research results and inspiration. Thanks to the team of authors for their efforts in experimental design, data analysis, and paper writing.

link: Wan, Weixiang, et al. "Development and Evaluation of Intelligent Medical Decision Support Systems." Academic Journal of Science and Technology 8.2 (2023): 22-25.

REFERENCES

[1] Grace M C , Zhou W , Jiang X , et al. Unsafe Exposure Analysis of Mobile In-App Advertisements ABSTRACT[C// Acm Conference on Security & Privacy inWireless & Mobile Networks.ACM2012.

[2] Tianbo, Song, Hu Weijun, Cai Jiangfeng, Liu Weijia, Yuan Quan, and He Kun. "Bio-inspired Swarm Intelligence: a Flocking Project With Group Object Recognition." In 2023 3rd International Conference on Consumer Electronics and Computer Engineering (ICCECE), pp. 834-837. IEEE, 2023.DOI: 10.1109/mce.2022.3206678

[3] Che, C., Liu, B., Li, S., Huang, J., & Hu, H. (2023). Deep learning for precise robot position prediction in logistics. Journal of Theory and Practice of Engineering Science, 3(10), 36-41.

[4] Liu, B., Zhao, X., Hu, H., Lin, Q., & Huang, J. (2023). Detection of Esophageal Cancer Lesions Based on CBAM Faster R-CNN. Journal of Theory and Practice of Engineering Science, 3(12), 36–42.

[5] Liu, Bo, et al. "Integration and Performance Analysis of Artificial Intelligence and Computer Vision Based on Deep Learning Algorithms." arXiv preprint arXiv:2312.12872 (2023).

[6] Li, Linxiao, et al. "Zero-resource knowledge-grounded dialogue generation." Advances in Neural Information Processing Systems 33 (2020): 8475-8485.

[7] Shen, Zepeng, et al. "The Application of Artificial Intelligence to The Bayesian Model Algorithm for Combining Genome Data." Academic Journal of Science and Technology 8.3 (2023): 132-135.

[8] Zong, Yanqi, et al. "Improvements and Challenges in StarCraft II Macro-Management A Study on the MSC Dataset." Journal of Theory and Practice of Engineering Science 3.12 (2023): 29-35.

[9] Zhang, Quan, et al. "Deep Learning Model Aids Breast Cancer Detection." Frontiers in Computing and Intelligent Systems 6.1 (2023): 99-102.

[10] Xu, Jingyu, et al. "Based on TPUGRAPHS Predicting Model Runtimes Using Graph Neural Networks." Frontiers in Computing and Intelligent Systems 6.1 (2023): 66-69.

[11] Liu, Yuxiang, et al. "Grasp and Inspection of Mechanical Parts based on Visual Image Recognition Technology." Journal of Theory and Practice of Engineering Science 3.12 (2023): 22-28.

[12] Zong, Yanqi, et al. “Improvements and Challenges in StarCraft II Macro-Management A Study on the MSC Dataset”. Journal of Theory and Practice of Engineering Science, vol. 3, no. 12, Dec. 2023, pp. 29-35, doi:10.53469/jtpes.2023.03(12).05.