International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161
Available at http://inpressco.com/category/ijcet

Research Article

A Framework to Predict the Advertise View Ability by using Machine Learning

Ms.Pooja Bhikaji Kamble and Dr.Nihar Ranjan

Department of Computer Engineering. JSPM Narhe Technical Campus ,Pune

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

Advertisements assume a crucial job in each industry and they help in the development of the business. The commercials that are distributed are not seen appropriately by the client in view of inadequate scrolling. By utilizing the scroll measurement process, we can foresee the view ability of the Advertisements dependent on the scrolling rate and and also predict the maximum viewed advertisements. Online advertisement has become a billion-dollar industry, and it continues developing. Advertisers endeavor to send marketing messages to draw in potential clients by means of realistic flag promotions on distributes website pages. Promoters are charged for each perspective on a page that conveys their presentation advertisements. Notwithstanding, ongoing investigations have found that the greater part of the advertisements are never appeared on clients' screens because of inefficient scrolling. Hence, advertisers waste a lot of cash on these promotions that don't bring any return on investment.

Keywords: *Computational Advertising, View-ability Prediction, User Behavior.*

Introduction

Online advertising advertising has emerged as one of the most renowned types of advertising. Studies show that publishing advertising has created earnings of over 98.2 billion in 2018. Internet advertising involves a distributor, who integrates promotions into his online substance, and an advertiser who provides advertisements to be published.Display ads can be seen in a wide range of formats and contain items such as text, images, flash, video and audio. In display advertising, an advertiser wages a publisher for space on web pages to display a banner during page views in order to impress the visitors who are interested in his products. Display ads can be seen in a wide range of formats and contain items such as text, images, flash, video and audio. In display advertising, an advertiser wages a publisher for space on web pages to display a banner during page views in order to impress the visitors who are interested in his products. A page view happens each time a web page is requested by a customer and displayed on internet. One-time display of an advertisement in a page view is called an ad impression, and it is considered the basic unit of advertisement delivery. Advertiser's wages for the ad impressions with the expectation that their advertisements will be viewed, clicked on, or converted by customers (e.g., the ad results in a purchase). Certainly, customers like to purchase products from the varieties that they recognize and trust. Display advertisements can make an expressive experience that gets customers surprised about varieties and creates some trust.To note this problem, another pricing model, which wages advertisements by the number of rendering imitations that a publisher has served, has become popular in the display advertising field. However, a modern study shows that more than half number of the imitations are actually not seen by customers because they do not scroll down a page enough to view the advertisements.

B. Motivation

Advertisements play a essential role in every enterprise and they help in the increase of the enterprise. The advertisements which are published are not regarded well by using the user just because of insufficient clicking. We don't have any platform that can provide us the standard web-page depth that is visited by each and every user that logs in your website.

C. Objectives

1. To develop the system for recording scroll depth.
2. To record data and use it as data set.
3. To each users scroll depth.
4. To arrange the web pages as per recorded values

Review of Literature

In the [1] work, user suggested, "Word representations: A simple and general method for semi-supervised learning". We use near state-of-the-art supervised baselines, and find that each of the three-word representations improves the accuracy of these baselines. We find further improvements by combining different word representations. The disadvantage, however is that accuracy might not be as high as a semisupervised method that includes task-specific information and that jointly learns the supervised and unsupervised tasks.In the [2] work, The objective of this paper is to present the design of personalized click prediction models. An essential part of these models is the development of new user-related features. We base our features on observations over a significant volume of search queries from a large number of users. Our observations suggest that user click behavior varies significantly with regard to their demographic background, such as age or gender. We investigate the click distribution for different users from various backgrounds and design a set of demographic features to model their group clicking patterns.Recognizing that there is still significant variability in demographic groups, we also investigate userspecific features.In the [3] work, author states, "On the importance of initialization and momentum in deep learning". Stochastic gradient descent with momentum. In this paper, we show that when stochastic gradient descent with momentum uses a well-designed random initialization and a particular type of slowly increasing schedule for the momentum parameter, it can train both DNNs and RNNs (on data sets with long-term dependencies) to levels of performance that were previously achievable only with Hessian-Free optimization. We find that both the initialization and the momentum are crucial since poorly initialized networks cannot be trained with momentum and well-initialized networks perform markedly worse when the momentum is absent or poorly tuned.In the [4] work, author states that, "The efficient back prop, in Neural networks: Tricks of the trade". The convergence of back-propagation learning is analysed so as to explain common phenomenon observed by practitioners. Many undesirable behaviours of back propagation can be avoided with tricks that are rarely exposed in serious technical publications. This paper gives some of those tricks, and offers explanations of why they work. Many authors have suggested that second-order optimization methods are advantageous for neural net training. It is shown that most "classical" secondorder methods are impractical for large neural networks. A few methods are proposed that do not have these limitations. In the [8] work, author states that, "The sequential click prediction for sponsored search with recurrent neural networks". Click prediction is one of the fundamental problems in sponsored search. Most of the existing studies took advantage of machine learning approaches to predict ad clicks for each event of ad view independently. However, as observed in the real-world sponsored search system, user's behavior on ads yield high dependency on how the user behaved along with the past time, especially in terms of what queries she submitted, what ads she clicked or ignored, and how long she spent on the landing pages of clicked ads, etc.In the [11] author present user engagement classes provide clear and interpret-able taxonomy of user engagement with online news, and are defined based on amount of time user spends on the page, proportion of the article user actually reads and the amount of interaction users performs with the comments.In the [12] author present Asking the advertisement of Ali search advertising as the research object, a feature processing method based on the pre-analysis of store and user data is put forward, and then the conversion rate is predicted with XGBoost (eXtreme Gradient Boosting). Experiments show that compared with other priori Feature Engineering, the proposed method can significantly improve the prediction results.

Proposed Methodology

This approach is based on machine learning so we work web page scroll depth and we are going to use EM
Maximization algorithm.
First, we are going to record each and every users scroll depth of web page that is going to be visited.
After that we are going to use the machine learning method that will be used to predict the standard web page depth of the page so that the advertisement can be arranged and profitability can be increased.
Advantages of Proposed System:

1. This system will use machine learning method to predict the web page depth.
2. This work on machine learning Expectation Maximization(EM) Algorithm.
3. This proposed system effectively able to record thescrolling percent of web page.
4. This proposed system accurately calculate the web pagedepth prediction.

A. Architecture
Explanation:
Machine Learning
In this step we are going to use the machine learning methodology for working on scroll depth measurement of web page and predict the standard web page depth of the page.
Project flow Steps:

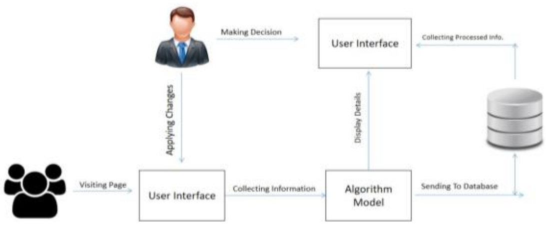

Fig. 1. Proposed System Architecture

1) The user will be register first.
2) Then after logging in the user can view web pages
3) while scrolling web page the scroll depth will be measured

4) The scroll depth and user details will be stored in database for use purpose
5) By using Machine learning algorithm we are going to calculate the standard web page depth.
6) And further changes will be made as per result.

B. Algorithms

1. Expectation Maximization (EM) Algorithm:

It can be used as the basis for unsupervised learning of clusters.

It can be used for the purpose of estimating the parameters of Hidden Markov Model (HMM).

Finding a maximum likelihood solution typically requires taking the derivatives of the likelihood function with respect to all the unknown values, the parameters and the latent variables, and simultaneously solving the resulting equations. The EM algorithm performs an expectation step (E-step) and a maximization step (M-step) Alternatively. Initially, a set of initial values of the parameters are considered. A set of observed data is given to the system that the observed data comes from our model.

C. Mathematical Model

1. Mathematical equation in EM Maximization:

$$\mathcal{N}(x|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2}}\frac{1}{|\Sigma|^{1/2}}exp\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) \quad ($$

Introduction of the Guassian Mixture Model The Guassian Mixture distribution The Guassian Mixture distribution is a linear superposition of Guassians:

$$p(x) = \Sigma_{k=1}^{K}\pi_k\mathcal{N}(x|\mu_k,\Sigma_k) \quad (2)$$

Subject to:

$$\Sigma_{k=1}^{K}\pi_k = 1 \quad (3)$$

Introduction of the Guassian Mixture Model The Guassian Mixture distribution

Introduction of the Guassian Mixture Model Now, for a Guassian Mixture Model, given the parameters: k, the number of Guassian components $\pi_1...\pi_k$, the mixture weights of the components $\mu_1...\mu_k$, the mean of each component

$\Sigma_1...\Sigma_k$, the variance of each component We can generate samples $s_1,s_2...s_n$ from the distribution.

Why do we need Guassian Mixture

The latent variable Given a Guassian Mixture model, we introduce K-dimensional binary random variable z which only one element z_k is euqal to 1 and the others are all 0.

z = (0,0,..,1,0,..0) (4)

So there are K possible states for z.And we let

p(z_k = 1) = π_k (5)

Result and Discussion

We will be getting the records of each and every users' web-page scrolling depth so that it can help us to predict the standard web-page depth for our particular website. The data will be recorded of each and every individual for performing operations. By implementing this project, we are providing a platform for various publishers or website owners to predict the content viewability of the page. Solving this issue can benefit online advertisers to allow them to invest more effectively in advertising and can benefit publishers to increase their revenue.

Conclusion

By implementing this project, we are providing a platform for various publishers or website owners to predict the content viewability of the page. Solving this issue can benefit online advertisers to allow them to invest more effectively in advertising and can benefit publishers to increase their revenue.

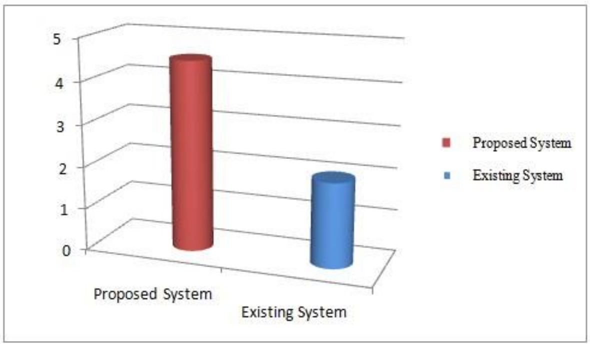

Fig. 2. Algorithms Comparison

Acknowledgment

The authors would like to thank the researchers as well as publishers for making their resources available and teachers for their guidance. We are thankful to the authorities of University of Pune and concern members of cPGCON 2019 conference, organized by, for their constant guidelines and support. We are also thankful to the reviewer for their valuable suggestions. We also thank the college authorities for providing the required infrastructure and support. Finally, we would like to extend a heartfelt gratitude to friends and family members.

References

[1]. I. Lunden, "Internet ad spend to reach 121b in 2014,"http://techcrunch.com/ 2014/04/07 /internet-ad-spend-to-reach-121bin-2014-23-of-537b-total-ad-spend-ad-tech-gives-display-aboost-oversearch/

[2]. Y. Chen and T. W. Yan, "Improving accuracy and cost of two-class and multi class probabilistic classifiers using ROC curves," in KDD'12, 2012, pp. 795–803.

[3]. W. Zhang, S. Yuan, and J. Wang, "Optimal real-time bidding for display advertising," in ACM SIGKDD'14, 2014, pp. 1077–1086.

[4]. W. Chen, D. He, T.-Y. Liu, T. Qin, Y. Tao, and L. Wang, "Generalized second price auction with probabilistic broad match," in ACM EC'14, 2014, pp. 39–56.

[5]. Google, "The importance of being seen", http://think.storage.googleapis.com/docs/the-importanceof-being-seen study.pdf.

[6]. M. Mareck, "Is online audience measurement coming of age?" Research World, vol. 2015, no. 51, pp. 16–19, 2015.

[7]. S.Flosi, G. Fulgoni, and A. Vollman, "if an advertisement runs online and no one sees it, is it still an ad?" Journal of Advertising Research, 2013.
[8]. H. Cheng, E. Manavoglu, Y. Cui, R. Zhang, and J. Mao, "Dynamic ad layout revenue optimization for display advertising," in Proceedings of the Sixth International Workshop on Data Mining for Online Advertising and Internet Economy, 2012, p. 9.
[9]. H. Weinreich, H. Obendorf, E. Herder, and M. Mayer, "Not quite the average: An empirical study of web use," ACM TWEB, vol. 2, no. 1, p. 5, 2008.
[10]. F. Manjoo, "You won't finish this article," Slate, 2013. Security, vol. 13, 2018.
[11]. D. Lagun and M. Lalmas, "Understanding user attention and engagement in online news reading," in Proceedings of the Ninth ACM International
[12]. Conference on Web Search and Data Mining. ACM, 2016, pp. 113–122.
[13]. Yue Sun and Guowei Yang, "Feature Engineering for Search Advertising Recognition" in 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)