A Survey paper on framework to predict the advertise view ability by using machine learning

MS.POOJA KAMBLE

ME Student, Department of Computer Engineering, Rajarshi Shahu School of Engineering and Research.
JSPM Narhe Technical Campus Pune.

DR.NIHAR RANJAN

Professor, Department of Computer Engineering, Rajarshi Shahu School of Engineering and Research.
JSPM Narhe Technical Campus Pune.

Abstract

Advertisements play a vital role in every industry and they help in the growth of the industry. The advertisements that are published are not viewed properly by the user just because of insufficient clicking. By using the scroll depth process, we are able to predict the view ability of the advertisement based on the scroll percentage and also predict the maximum viewed advertisements. Online display advertising has become a billion-dollar industry, and it keeps growing. Advertisers attempt to send marketing messages to attract potential customers via graphic banner ads on publishers' web pages. Advertisers are charged for each view of a page that delivers their display ads. However, recent studies have discovered that more than half of the ads are never shown on users' screens due to inefficient scrolling. Thus, advertisers waste a great amount of money on these ads that do not bring any return on investment.

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

1. INTRODUCTION

Online display advertising has arisen as one of the most famous forms of advertising. Studies show that publishing advertising has generated earnings of over $98.2 billion in 2018. Internet advertising involves a publisher, who integrates advertisements into his online content, and an advertiser, who provides advertisements to be published. Displays 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. Advertisers

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). Old fashioned display advertisement compensation is mainly based on customer clicks and changes, because they bring direct income to the advertisers. Much research has been done for probabilistic analysis on the click rate and conversion rate, bid optimization and public sales of companies and their products. 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 is actually not seen by customers because they do not scroll down a page enough to view the advertisements. Low view ability leads to inefficient brand advertising. Therefore, a new pricing model is growing: pricing advertisement by the number of imitations that can be seen by a customer, instead of just being used. This eliminates the stress of the advertisers, who concern about paying for advertisements that were served but not viewed by customers. Unsurprisingly, the advertisements located at different page depths have different probabilities of being seen by a customer. Therefore, it is mandatory to judge the probability that an advertisement at a known page depth will be viewed on a customer's screen, and thus be measured as seen. The vertical page depth that a customer scrolls to is defined as the scroll depth.

1.1 Motivation

- Advertisements play a vital role in every industry and they help in the growth of the industry. The advertisements that are published are not viewed properly by 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.

1.2 Problem Statement

- The most important source of income for content publishers or website owners is the marketing of the products or things. As per the Google's payment models the content

owners are paid if and only if the advertisement is viewed by the user. We do not have any model that predicts the depth where the advertisement can be viewed by the user.

- Solving this issue can benefit online advertisers to allow them to invest more effectively in advertising and can benefit publishers to increase their revenue?

2. RELATED WORK

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 semi-supervised 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 user-specific 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 datasets 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" second-order 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. Inspired by these observations, we introduce a novel framework based on Recurrent Neural Networks (RNN). Compared to traditional methods, this framework directly models the dependency on user's sequential behavior into the click prediction process through the recurrent structure in RNN. Large scale evaluations on the click-through logs from a commercial search engine demonstrate that our approach can significantly improve the click prediction accuracy, compared to the sequence-independent approaches.

Chen et al. [9] introduced a factorization model to predict the advertisement with search results based on the click position. This prediction is done for a given position and pair of query advertisements. It does not consider the web user as a factor

Liu et al. [10] proposed dwell time data with Waybill distributions and demonstrate the possibility of predicting webpage dwell time distribution from page-level features. Dwell time data specifies time spend by the user in the same position in a web page.

Yi et al. [11] predict dwell time through Support Vector Regression, using the context of the webpage as features. However, both methods do not consider individual user characteristics, an important factor of scrolling prediction.

Chong Wang et at [12] proposes two probabilistic latent class models (PLC) with constant membership and dynamic memberships that predict the view ability of any given scroll depth for a user-page pair. Using a real-life dataset from a large publisher, the experiments demonstrate that the models outperform comparison systems.

M. Mareck [13] finds new pricing model for pricing ads by the number of impressions that can be viewed by a user not for unseen by the user. This model avoids the payments for served in the page but not unseen by the use.

SwathiVoddi et al. [14] recommends a proficient solution using PLC and expectation maximization algorithm for anticipating the visibility prediction of ads taking in view the scrolls, number of clicks per user and web page pair.

JurgitaRaudeli et al [15] carrying out an experiment to examine the assessment factors of the effectiveness of advertising campaigns on social media networks and their applicability to various types of organizations, taking into account the specifics of the activities these organizations operate; and evaluating the interconnectivity of the assessment variables of advertising campaigns and their impact on the performance of an organization. During the study, out of 39 primary assessment factors, eight primary factors that influence the effectiveness of advertising campaigns on social media networks were identified: sales, content reach, and traffic to website, impressions, frequency, relevance score, leads and audience growth.

Malar vizhiet al [16] improves the credit ability of the online advertisers and it leads to more investment on advertising by the

growing and upcoming industries to earn more money using two probabilistic latent class models.

Eugenere Agichtein et al [17] collects scrolling behaviour and considers scrolling behaviour as an implicit indicator of user interests in order to measure the quality of webpage design and its content. The author proposed predictive model of the scrolling behaviour for any user on any based on historic information. The focus is to make prediction before the user behaviour occurs, rather than observation or measurement.

Wang et al[18], learns user's click behaviour from server logs in order to predict if a user will click and ad shown for the query. The authors use features extracted from the queries to represent the user search intent. In this research proposal, search queries, which can explicitly reflect user interest is not available.

Olivier Chappelle et al [19] are done on the advertiser side, based on high dimensional features about user's profiles, ad campaigns and impression content. On the other hand, such data is not available at the publisher side. Therefore, these existing techniques of predicting click Behaviour cannot be readily used to predicting scrolling behaviour at the publisher side.

Kim et al [20] presents regression method to estimate the parameters of the Gamma distribution of click dwell time. But its predict dwell time at a specific depth in page and depth-level dwell time prediction is more challenging than page-level dwell time prediction.

Abu Tair and El-Halees gave a comprehensive case study from the higher education stage [21]. The main purpose of their study is to show how useful data mining can be in the educational domain, their research discovered many kinds of knowledge from the graduate student dataset using different educational data mining techniques, this acquired knowledge included classification, clusters, association rule, and outlier detection.

M. Sukanya, S. Biruntha, S. Karthik and T. Kalaikumaran applied the Bayesian classification technique on existing higher education students [22]. The main goal of their study is to predict the number of upcoming students in the next year based on the number of enrolled students in the previous years. This study helps decision makers to manage the resources and staffs they need to administer the outcomes of a student. This study helps also the teachers to identify at early stage the students that need more attention to facilitate taking the correct action at the suitable time to reduce the failure in the academic approach and improve the student's academic performance.

Md. Hedayetul Islam Shovon and MahfuzaHaque[23] implemented k-means cluster algorithm. The main goal of their study is to improve the quality of the education by dividing the students into groups according to their characteristics using the application which have implemented.

Er. RimmyChuchra, M. tech [24] gave a case study from the higher education university. Their study was based on applying neural networks on the existing student dataset from the university database to build decision trees that can be used to evaluate the performance of students.

Brijesh Kumar Bhardwaj and Saurabh Pal [25] applied Baysian classification on the student database from the higher education

stage. This study aimed at identifying those students who needed more attention to reduce the drop out ratio and take action at a right time.

3. PROPOSED METHOD

To classify the web-page page depth prediction we are going to use Expectation Maximization (EM) Algorithm. The EM algorithm is widely used to solve the maximum-likelihood parameter estimation problem.

- 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).
- It can be used for classifying the values of depth variables.
- 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.
- The next step is known as "Expectation"-step or E-step. In this step, we use the observed data in order to compare the variables.
- The next step is known as "Maximization"-step or M-step. In this step, we use the complete data to classify the data.
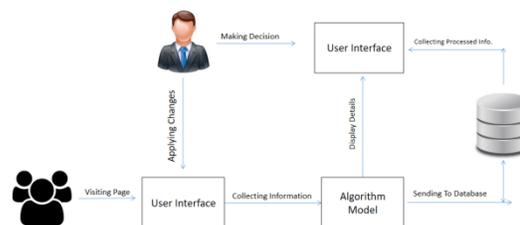
Architecture

Fig.2 System Architecture

4. EXPECTED RESULT

We will be getting the records of each and every user's 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.

5. CONCLUSION

By implementing this project, we are providing a platform for various publishers or website owners to predict the content view ability 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?

6. References

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

[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] Y. Chen and T. W. Yan, "Position-normalized click prediction in search advertising," in KDD'12, 2012, pp. 795–803.

[12] C. Liu, R. W. White, and S. Dumais, "Understanding web browsing behaviors through weibull analysis of dwell time," in ACM SIGIR'10, 2010, pp. 379–386.

[13] X. Yi, L. Hong, E. Zhong, N. N. Liu, and S. Rajan, "Beyond clicks: dwell time for personalization," in Recsys'15, 2014, pp. 113–120.

[14] Chong Wang, AchirKalra, Li Zhou, CristianBorcea, Member, IEEE and Yi Chen, "Probabilistic Models For Ad View ability Prediction On The Web", Article in IEEE Transactions on Knowledge and Data Engineering • May 2017, DOI: 10.1109/TKDE.2017.2705688

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

[16] SwathiVoddi, Phani Sri Koritala, PriyaDarshiniMuchu , "Anticipating Advertisement Visibility using Prospective Representation", InternationalResearch Journal of Trend in Scientific and Development, | Volume – 2 | Issue – 3 | Mar-Apr 2018

[17] JurgitaRaudeli, Vida, Manuela, "Evaluation of Advertising Campaigns on Social Media Networks", Sustainability 2018, 10, 973; doi:10.3390/su10040973

[18] Chein-Jen Wang and HsinHsi Chen, "Learning user behaviors for advertisements click prediction", International ACM SIGIR Conference on Research and Development in Information Retrieval, Pages 1-6

[19] Olivier Chapelle, Eren and Romer Rosales, " Simple and Scalable response prediction for display advertising", ACM Transaction on Intelligent Systems and Technology, 5(4):61:-161:34, December -1

[20] Kim, Hassan, Ryen and Imed, "Modeling dwell time to predict click-level satisfaction", In proceedings of the 7th ACM International conference on Web Search and Data mining, WSDM'14, Pages 193-204

[21] A.M. El-Halees, and M.M. Abu Tair, "Mining educational data to improve students' performance: A case study," International Journal of Information and Communication Technology Research, 2011, pp. 140-146.

[22] S.KarthikM.Sukanya, S.Biruntha and T.Kalaikumaran, "Mining Data mining: Performance improvement in education sector using classification and clustering algorithm," ICCCE In Proceedings of the International Conference on Computing and Control Engineering, 2012.

[23] MahfuzaHaque Md. Hedayetul Islam Shovon. "Prediction of student academic performance by an application of k-means clustering algorithm". International Journal of Advanced Research in Computer Science and Software Engineering, 2(7):353–355, July2012.

[24] M. tech Er. RimmyChuchra. "Use of data mining techniques for the evaluation of student performance: a case study". International Journal of Computer Science and Management Research, 1(3):425–433, October 2012.

[25] Brijesh Kumar Bhardwaj and Saurabh Pal. Data mining:" A prediction for performance improvement using classification". (IJCSIS) International Journal of Computer Science and Information Security, 9(4), April,2011.

[26] Nihar M. Ranjan & Rajesh S. Prasad "Automatic text classification using BPLion-neural network and semantic word processing" ISSN:1368-2199 (Print) 1743-131X (Online) Journal homepage: http://www.tandfonline.com/loi/yims.

[27] Nihar M. Ranjan & Rajesh S. Prasad "LFNN: Lion fuzzy neural network-based evolutionary model for text classification using context and sense based features. Applied Soft Computing" Applied Soft Computing 8-4946(18)30404-6(2018).

[28] Prof Nihar Ranjan , Dr. R. S.Prasad "IDENTIFICATION IN TEXT MINING FOR USED IN FORENSICS."