Webpage Depth-level Dwell Time Prediction

Chong Wang
Information Systems Dept.
New Jersey Institute of Technology
Newark, NJ 07102, USA
cw87@njit.edu

Achir Kalra*
Forbes Media
499 Washington Blvd
Jersey City, NJ 07310
akalra@forbes.com

Cristian Borcea
Computer Science Dept.
New Jersey Institute of Technology
Newark, NJ 07102, USA
borcea@njit.edu

Yi Chen†
Martin Tuchman School of Management
New Jersey Institute of Technology
Newark, NJ 07102, USA
yi.chen@njit.edu

ABSTRACT

The amount of time spent by users at specific page depths within webpages, called dwell time, can be used by web publishers to decide where to place online ads and what type of ads to place at different depths within a webpage. This paper presents a model to predict the dwell time for a given <user, webpage, depth> triplet based on historic data collected by publishers. Dwell time prediction is difficult due to user behavior variability and data sparsity. We adopt the Factorization Machines model because it is able to capture the interaction between users and webpages, overcome the data sparsity issue, and provide flexibility to add auxiliary information such as the visible area of a user's browser. Experimental results using data from a large web publisher demonstrate that our model outperforms deterministic and regression-based comparison models.

Keywords

Computational Advertising; User Behavior; Data Mining

1. INTRODUCTION

Online display advertising provides many benefits that traditional marketing channels do not, such as fast brand building and effective targeting. In display advertising, an advertiser pays a publisher for space on webpages to display a banner during page views in order to attract visitors that are interested in its products. A *page view* happens each time a webpage is requested by a user and displayed in a browser. One display of an ad in a page view is called an *ad impression*, the basic unit of ad delivery.

Currently, there are two main ad pricing models, pay-by-action and pay-by-impression. In pay-by-action, advertisers are charged when the impressions are clicked or converted. But the rate of click or conversion are often very low. Also, some advertisers, e.g. car vendors, do not expect users to purchase their products through ads. They just want to increase their brand awareness and make more users be aware of their logos or products. In pay-by-impression, advertisers have to pay for the impressions served (i.e., sent to the users' browsers) but not *viewed* by users who do not spend time at the page depth where the ads are placed [1]. To solve this issue, a new model is emerging: pricing ads by the number of impressions viewed by users, instead of just being served. This is attractive for advertisers, who want to prevent investment waste. The Interactive Advertising Bureau (IAB) defines a *viewable impression* as one that is at least 50% shown on the screen for at least one second. Advertisers may require various visible ad areas and/or time durations. They can thus specify viewability requirements in guaranteed delivery contracts with publishers. Predicting the likelihood of an ad being viewed can be very helpful in many applications: In guaranteed delivery, according to different viewability requirements, publishers can determine which ad to be served by predicting its viewability, in order to maximize the revenue. In real-time bidding, advertisers can decide bidding price based on the predicted viewability. Therefore, ad viewability prediction is essential to fulfill the marketing requirements and thus maximize the return on investment for advertisers as well as boost the advertising revenue for publishers.

However, the only existing work on viewability prediction is done by Wang et al. [6]. They propose a probabilistic latent class model that predicts the probability that a user scrolls to a page depth where an ad may be placed, but not the dwell time. Also, all existing studies on dwell time prediction focus on the time a user spends on an entire page, not a specific page depth. Liu et al. [4] predict the Weibull distributions of page-level dwell time using regression trees with features, e.g., keywords, page size. Yi et al. [8] use

*The author is also with the Computer Science Department, New Jersey Institute of Technology, USA.
†Corresponding Author.

CIKM '16, October 24–28, 2016, Indianapolis, IN, USA.

DOI: http://dx.doi.org/XXXX.XXXX

support vector regression to predict page-level dwell time with features such as content length, topical category, and device. Kim et al. [2] present a regression model to estimate the Gamma distributions of the time that a user spends on a clicked result. Xu et al. [7] propose a webpage re-ranking algorithm by estimating page-level dwell time. Yin et al. [9] state that viewing textual items is such a casual behavior that people may terminate the viewing process at any time. The authors develop a model to estimate user preferences according to the item-level (page-level) dwell time.

In contrast, our work is the first to predict dwell time at a specific depth in a page view. Working at a finer granularity, depth-level dwell time prediction is more challenging than page-level dwell time prediction. This open problem is non-trivial due to the variability of user behavior and data sparsity, i.e., most users read only a few webpages, while a webpage is visited by a small subset of users. It is also difficult to explicitly model user interests as well as the characteristics of entire pages and depths. Thus, we explore a machine learning model to predict the dwell time at a page depth where an ad is placed, i.e., the time that the ad is shown on screen. The proposed method can also be applied to predict the dwell time of any items on a page. We adopt the Factorization Machines (FM) model because it is able to capture the interaction between input features, overcome the data sparsity issue, and provide flexibility to add auxiliary information. Our FM models consider basic factors (i.e., user, page, and page depth) and auxiliary information such as context features. We determined experimentally that viewport (i.e., the visible area of a user browser) is the most important context feature. We evaluated our model using real-data from a Forbes Media, a large web publisher. The experimental results demonstrate that our model outperforms the comparison models.

2. DEPTH DWELL TIME PREDICTION

Problem Definition. *Given a page view, i.e., a user u and a webpage a, the goal is to predict the dwell time of a given page depth X, i.e., the time duration that X is shown on the screen. The dwell time of X is denoted as T_{uaX}.*

2.1 Dataset

A large web publisher (i.e., Forbes Media) provides user browsing logs collected from real website visits in one week and webpage metadata. The dataset contains 2 million page views. For each page view, it records the user id, page url, state-level user geo location, user agent, and browsing events, e.g. the user opened/left/read the page. Each event stores the event time stamp and the page depths where the top and bottom of the user screen are. Once a user scrolls to a page depth and stays for one second, an event is recorded. The page depth is represented as the percentage of the page. The reason that we adopted page percentage rather than pixels is because it provides a relative measure independent of device screen size. If a user reads 50% of a page on a mobile device, while another user reads 50% of the same page on a desktop, it is assumed that they read the same content.

Each event has a time stamp so that the time that a user spent on a part of a page can be calculated. To infer the current part of a page that a user is looking at, the user log also records the page depths at which the first and the last rows of pixels of the screen are. Thus, we are able to infer the part of the page to which the user scrolls and how long the user stayed at that part of the page. Therefore, the dwell time at a page depth can be easily calculated from the information provided by the user log.

2.2 Model

It is intuitive that the dwell time of a page depth is highly related to the user's interests and reading habits, the topic of the article in the page, the design at that page depth, etc. For instance, some users tend to stay longer on pages, while some are less patient. A viral content may attract most users to scroll deep on the page and spend a long time on the whole page. Page depths with important topic sentences may keep most users longer on them. Thus, the characteristics of individual users, webpages, and page depths should be taken into account for depth-level dwell time prediction. More importantly, the interactions of these three factors must be modeled so that their joint effect is captured: 1) The interaction of users and pages captures a user's interest in a page. 2) The interaction of users and page depths can reflect individual users' browsing habits. For example, some users read entire pages carefully, but some only read the upper half. 3) The interaction of pages and depths models the design of individual pages at individual page depths. For example, pages that have a picture at a depth may receive relatively short dwell time at that depth because people usually can understand a picture more quickly than text. However, it is non-trivial to explicitly model user interests, page characteristics, the attractiveness of page depths, and their interactions. Also, although implicit feedback, e.g. reading dwell time, is more abundant than explicit feedback, e.g. ratings, it often has higher variability [9], which makes prediction more challenging.

We adopt Factorization machines (FM) [5], which are a generic approach that combines the high-prediction accuracy of factorization models with the flexibility of feature engineering. The reason that we adopt the FM model is that it can capture the interaction of multiple inter-related factors, overcome the data sparsity, and provide the flexibility to add auxiliary information.

According to the problem definition, the basic FM model requires three factors: user, page, and page depth. The input is derived from the user-page-depth matrix built from the user logs: In the basic form of depth-level dwell time prediction, we have a 3-dimensional cube containing n_u users, n_a pages, and n_d page depths. Thus, each dwell time is associated with a unique triplet <user, page, depth>. Such a 3D matrix can be converted into a list of $(n_u + n_a + n_d)$ rows. The target variable for each row corresponds to an observed dwell time represented by the triplet. N training page views lead to $N \cdot 100$ rows, as each page view contains 100 observed dwell time values (one for each percent from 1% to 100% page depth). This input is similar to what is prepared for regressions. However, regressions would not work well because the data is very sparse and they are unable to capture the interaction between the input variables.

The basic idea of FM is to model each target variable as a linear combination of interactions between input variables. Formally, it is defined as following.

$$\hat{y}(\mathbf{x}) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \langle \mathbf{v_i}, \mathbf{v_j} \rangle \, x_i x_j \qquad (1)$$

where, $\hat{y}(\mathbf{x})$ is the prediction outcome given an input $\mathbf{x}$. w_0 is a global bias, i.e., the overall average depth-level dwell

time. $\sum_{i=1}^{n} w_i x_i$ is the bias of individual input variables. For example, some users would like to read more carefully than others; some pages can attract users to spend more time on them; some page depths, e.g., very bottom of a page, usually receive little dwell time. The first two terms are the same as in linear regression. The third term captures the sparse interaction between each pair of input variables.

Unlike standard regression models which model the weight of each interaction by a real number w_{ij}, the FM model uses a factorized parametrization to capture the interaction effect (Eq. 2). Such low-rank interaction allows the FM model to estimate reliable parameters even in sparse data.

$$\langle \mathbf{v_i}, \mathbf{v_j} \rangle = \sum_{k=1}^{K} v_{ik} v_{jk} \quad (2)$$

The basic FM model works with only three factors: user, page, and depth. However, context information can also help improve the prediction performance. Thus, we identify two context features, *viewport* (i.e., the part of a user browser visible on the screen) and *local day of the week*, which are intuitively related to user reading behavior. The viewport indicates the device utilized by the user (e.g., a mobile device usually have a much smaller visible browser area than a desktop) and can directly determine the user experience. Specifically, one viewport value consists of the height and the width of a browser, e.g., 1855×1107. To reduce sparsity, both heights and widths are put into buckets with size 100 pixels. For instance, 1855×1107 can be discretized into 18×11. The local day of the week, expected to reflect if users are working, is inferred from the GMT time and user geo provided in the user log.

In addition, although in theory user demographics and page attributes are already considered in the latent user and page dimensions, incorporating these additional sources of information as features may further improve the prediction accuracy in some applications [3]. For user demographics, we consider user geo locations because this is the only explicit feature about users that can be easily obtained by publishers. User geo, inferred from IPs, may reflect a user's interests and education, and it may determine the user's network condition. Specifically, geo is the country name if the user is outside USA or a state name if she is within USA. For page attributes, we consider article length and channel. Article length is represented by the word count of the article in the page, and it has been proven to be a significant factor impacting page-level dwell time [8]. Article lengths are put into buckets so that there are a limited number of possible states. The channel of the article in a page is its topical category on the publisher's website, e.g., finance and lifestyle. A channel can be a high-level topic label of a page.

Context and auxiliary features, i.e., user geo and page attributes, can be used to extend the basic FM model. In Section 3.4, we compare the prediction performance of different combinations of auxiliary features.

3. EXPERIMENTAL EVALUATION

3.1 Experiment Datasets

A one-week user log, collected as described in Section 2.1, is split into three sets of training and testing data. The experimental results are reported by taking the average over the sets. On average, the training and test data contain 150K+ and 20K+ page views, respectively. The training/test data consist of all depths of all training/test page views.

3.2 Comparison Models

GlobalAverage: This model is used in two ways. In Section 3.5, it computes the average dwell time of each page depth X in all training page views. If a user did not scroll to X, its dwell time in the page view is zero. In Section 3.6, it computes the fraction of training page depths whose dwell time is no less than the required dwell time. The 100 constant numbers obtained are used to make a deterministic prediction for the page depth.

ChannelAverage: It is similar to GlobalAverage, but it computes the average dwell time of each depth X of the same page channel (rather than all training page views).

Regression: We built two regression models. 1) *Regress_bc* is developed based on an existing work on page dwell time prediction [8]. To apply it to depth-level prediction, one more feature, i.e., page depth, is added. 2) *Regress_view+dep* is developed based on the finding in Section 3.4 that shows the viewport to be the best feature for improving prediction. Thus, it has only two input features: viewport and depth. In the viewability prediction test, logistic regression with the same features is adopted. They outputs the probability that the dwell time of X is at least a certain seconds.

3.3 Metrics

The metrics we adopt are Root-Mean-Square Deviation (RMSD) and Logistic Loss. Both serve to aggregate the magnitudes of the errors in predictions for various times into a single measure of the predictive power of a method. Thus, for both metrics, lower values are better.

RMSD: $RMSD = \sqrt{\frac{\sum_{j=1}^{N}\sum_{i=1}^{100}(\hat{y}_{ij}-y_{ij})^2}{N \cdot 100}}$ measures the differences between the values predicted, $\hat{y}_i$, and the values observed, y_i. N is the number of test page views. The second sum accumulates the errors at all 100 page depths in the ith page view. y_{ij} is the actual dwell time at the jth page depth in the ith page view.

Logistic Loss: It is widely used in probabilistic classification. Compared to the RMSD, it penalizes a method more for being both confident and wrong.

$$logloss = -\frac{1}{N \cdot 100} \sum_{i=1}^{N} \sum_{j=1}^{100} [y_{ij} \log(\hat{y}_{ij}) + (1 - y_{ij}) \log(1 - \hat{y}_{ij})]$$

3.4 Comparison of Feature Combinations

Table 1: RMSD Comparison of Auxiliary Features

Approaches	K=10	K=20	K=30
FM	11.4667	11.5501	11.5925
FM (viewport)	11.141	**11.0309**	11.172
FM (dow)	11.4655	11.5331	11.5728
FM (geo)	11.5563	11.6735	11.7201
FM (length)	11.7064	11.727	11.8084
FM (channel)	11.7502	11.8827	11.9211
FM (viewport+dow)	11.5152	11.3235	11.3763
FM (viewport+geo)	11.0318	11.0767	11.2299
FM (viewport+length)	11.3026	11.5273	11.5666
FM (viewport+channel)	11.5319	11.3474	11.5725
FM (all five)	11.4084	11.7641	11.7385

We add context and auxiliary features into the basic FM model in order to evaluate the effect of different combinations. The results are presented in Table 1. The first row is the basic FM model. We vary the dimension of the 2-way interactions, K, which is the length of the latent vector v for each variable (Eq. 2).

The results show that viewport is the most significant context feature. Intuitively, viewport indicates the type of device, which influences reading experience and thus the way users engage with webpages. Local day of the week, denoted as "dow", does not improve the basic FM as much as viewport does. The three explicit user and page attributes do not enhance the performance of the basic FM. The possible reason is that the granularities of user geo and channels are too coarse. Learning latent features for each channel and geo cannot specifically capture the characteristics of individual users and pages. Article length may not play a key role at depth-level, as the text length in a screen is determined by the viewport size, not the length of the article. Also, increasing K does not always lead to performance improvement. Longer latent feature vectors may fit the data better, while they may cause overfitting. As the Bias-Variance trade-off, the optimal K can be obtained by cross-validation.

3.5 Depth-level Dwell Time Prediction

Table 2: Depth Dwell Time Prediction Comparison

Approaches	RMSD
GlobalAverage	13.8346
ChannelAverage	13.8219
Regress_bc	14.1009
Regress_view+dep	13.8301
FM (viewport;K=20)	11.0309

We compare the best model obtained from the previous experiment, i.e., FM (viewport) with K=20, with the baselines. All models are applied to predict the exact dwell time of each page depth in test page views. The results in Table 2 demonstrate that the FM model significantly outperforms the baselines. This is because it is able to overcome sparsity and capture pairwise interactions between users and pages. The RMSDs of GlobalAverage and ChannelAverage are similar. This suggests that controlling the channel variable does not help with performance. Regress_bc has the highest RMSD, which indicates that methods for page-level dwell time prediction cannot be easily applied to depth-level prediction. With only the combined viewport/depth feature, Regress_view+dep does not obtain a better prediction outcome than simple averaging.

3.6 Viewability Prediction

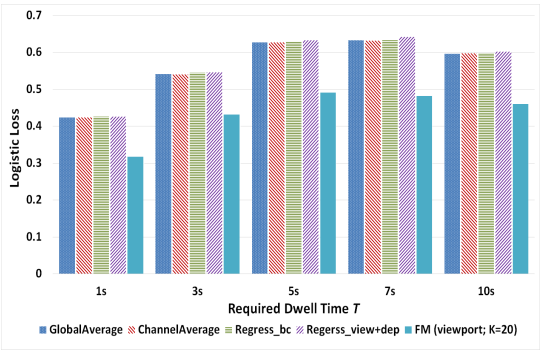

Figure 1: Viewability Prediction Comparison

We vary the dwell time threshold of a viewable impression from 1s (IAB standard) to 10s. For each page depth in the dataset, its target variable is 1 if its dwell time is at least T seconds; otherwise 0. In this way, the prediction problem is converted from regression to classification. The prediction outcome of each test page depth is the probability that its dwell time is at least T seconds.

Figure 1 shows that the FM model clearly outperforms the baselines. We also notice that the FM model achieves the best performance at the two ends (1s and 10s). Given a page depth, it is more challenging to predict if the dwell time is at least 5s. The reason is that the number of page depths with dwell time at least 5s and the number of page depths with dwell time less than 5s are very close (about 50%). In contrast, there are about 70% page depths whose dwell time is at least 1s. Similar to the depth-level dwell time prediction, GlobelAverage and ChannelAverage have very similar performance. Also, the Regress_bc and the Regress_view+dep are slightly worse than simple averaging.

4. CONCLUSIONS

Web publishers and advertisers are interested to predict how much time a user spends at different places in a webpage in order to maximize their profit and return on investment. This paper presents a model based on Factorization Machines to predict webpage depth-level dwell time for a page view. Using real-world data, both page depth-level dwell time and viewability prediction experiments consistently show our model outperforms the comparison models.

Acknowledgement

This work is partially supported by NSF under grants No. CAREER IIS-1322406, CNS 1409523, and DGE 1565478, by a Google Research Award, and by an endowment from the Leir Charitable Foundations. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

5. REFERENCES

[1] Google. The importance of being seen. https://think.storage.googleapis.com/docs/the-importance-of-being-seen_study.pdf,2014.

[2] Y. Kim, A. Hassan, R. W. White, and I. Zitouni. Modeling dwell time to predict click-level satisfaction. In *WSDM'14*, pages 193–202. ACM, 2014.

[3] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, (8):30–37, 2009.

[4] C. Liu, R. W. White, and S. Dumais. Understanding web browsing behaviors through weibull analysis of dwell time. In *SIGIR'10*, pages 379–386. ACM, 2010.

[5] S. Rendle. Factorization machines with libfm. *TIST*, 3(3):57, 2012.

[6] C. Wang, A. Kalra, C. Borcea, and Y. Chen. Viewability prediction for online display ads. In *CIKM'15*, pages 413–422. ACM, 2015.

[7] S. Xu, H. Jiang, and F. C.-M. Lau. Mining user dwell time for personalized web search re-ranking. In *IJCAI'11*, volume 22, page 2367, 2011.

[8] X. Yi, L. Hong, E. Zhong, N. N. Liu, and S. Rajan. Beyond clicks: dwell time for personalization. In *RecSys'14*, pages 113–120. ACM, 2014.

[9] P. Yin, P. Luo, W.-C. Lee, and M. Wang. Silence is also evidence: interpreting dwell time for recommendation from psychological perspective. In *KDD'13*, pages 989–997. ACM, 2013.