WimNet: Vision Search for Web Logs

Sungchul Kim, Sana Malik, Nedim Lipka, and Eunyee Koh
Adobe Research
San Jose, CA
{sukim, sanmalik, lipka, eunyee}@adobe.com

ABSTRACT

With the growing popularity of mobile devices, user web logs are more heterogeneous than ever, across an increased number of devices and websites. As a result, identifying users with similar usage patterns within these large sets of web logs is increasingly challenging and critical for personalization and user experience in many areas, from recommender systems to digital marketing.

In this work, we explore the use of visual search for top-k user retrieval based on similar user behavior. We introduce a convolution neural network (WimNet) that learns latent representation from a set of web logs represented as images. Specifically, it contains two convolution layers take row- and column-wise convolutions to capture user behavior across multiple devices and websites and learns latent representation and reconstructs a transition matrix between user activities of given web logs. To evaluate our method, we conduct conventional top-k retrieval task on the simulated dataset, and the preliminary analysis results suggest that our method produces more accurate and robust results regardless of the complexity of query log.

Keywords

User behavior; Top-k retrieval; CNN

1. INTRODUCTION

Fueled by the rapid growth of mobile devices, web logs contain varied user behaviors across multiple devices. As a result, identifying users with similar usage patterns within these large sets of web logs is increasingly challenging and critical for personalization and user experience in many areas, from recommender systems to digital marketing. Prior work on analyzing user behaviors has been done to enhance the performance of prediction models[3, 1, 5, 6]; however, most of these are dependent on the structure of the data itself, and do not provide representation of web logs for general purposes, including top-k retrieval. Many of these methods focus on finding similarities based on the timing between events (e.g., finding similar patterns among users), but few focus on the attributes about the events, such as which device was used or which website was visited. Although some work has been done towards accounting for these attributes at a user-level [2], there are intricacies which remain unsolved at the event-level of detail.

WWW 2017, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3054256

Figure 1: A sample visualization of user web logs where the x-axis is time and each colored (non-black) pixel represents a unit of activity (visit to a website).

In this work, we introduce a way to visualize web logs as an image (Figure 1) which encodes time, frequency, and two event attributes (device and website). Specifically, the events are colored by website and the pixel's brightness indicates frequency (where a single visit is less bright). The user's timeline is faceted into rows, where each row represents a user's device. For top-k retrieval of web logs, we suggest a convolution neural network that learns latent representation from a set of web log images. Specifically, it contains two convolution layers which take row- and column-wise convolutions to capture user behavior across multiple devices and websites. Though we use the example of web logs faceted based on device type, it is important to note that our approach can be extended for web logs faceted by any categorical attribute, such as platform, or location.

To evaluate our method, we conduct a preliminary comparison of our top-k method versus other conventional top-k retrieval methods on a simulated dataset. The results suggest that our method produces more accurate and robust results regardless of the complexity of query log. The baselines cannot accurately capture either the pattern of device transition or per-site visiting pattern.

2. METHOD AND MATERIAL

2.1 Dataset

We use a simulated dataset based on the visiting pattern of real-world web logs. It contains a set of user activity lists, where each user activity is a triplet, (t, d, v)[1]. Specifically, we assume that there are 4 devices (Desktop, mobile device, tablet, and game console) and 10 websites. For each user activity list, we start with sampling the initial device, web site, and visit frequency from two multinomial distributions and uniform distribution, respectively, and it is used as the

[1] t is a timestamp (a day), d is a device, and v is a website

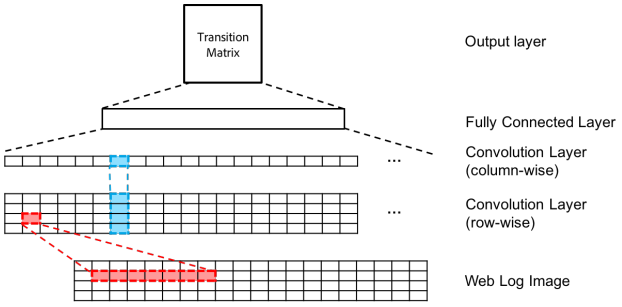

Figure 2: CNN architecture for learning latent representation of web logs

first visit log in the web log image. After that, for each timepoint (x-axis), we sample whether this user switched the device or not from bernoulli distribution, and re-sample the above three variables if it occurs, or sample the frequency only otherwise. Finally, we generate 1,000 web log images, and Figure 1 shows an example of a user's web log image using 3 devices while visiting several different websites with some noise.

2.2 WimNet

Based on the set of web log images, we learn a convolutional neural network (Figure 2), which consists of four layers including fully connected layer for representation learning. Its objective is to reconstruct transition matrix over user activities (a pair of device and website). Specifically, the row-wise convolution layer takes convolutions that capture the pattern of websites visited by a single device for a fixed amount of time (e.g., 1 week = 7 pixels). The column-wise convolution layer takes convolutions that capture the pattern of device transitions at each timepoint. Both of convolution layer use the rectified linear unit (ReLU) for activation function. Then, the fully-connected layer extracts latent representation from the convolution layer via a sigmoid function, $f(x) = 1/(1 + e^{-x})$, and we use it for top-k retrieval based on user behavior. In this work, the dimension of the latent representation set as the cardinality of user actions. Finally, from the previous layers, the output layer extracts transition matrix over user activity by using a sigmoid function.

3. ANALYSIS RESULTS

To evaluate our representation, we conduct a series of top-k retrieval. For comparison, we use four different approaches; 1) L1-distance between raw images (RAW), 2) Euclidean distance between two vectors of color histogram (HIST), 3) SIFT[4], and 4) polar distance between two subsequent sets of user activity (CHI)[6]. We posted source files and detailed results and at this link[2].

According to the result (Figure 3-(A)), when a query is simple, most of the approaches perform well except for SIFT. It can be contributed that SIFT is originally designed for detecting scale-invariant objects from regular images. However, web log images have low-resolution, and each pixel should be precisely considered because each pixel does have information of user activity.

When we use more complex query (Figure 3-(B)), RAW cannot produce accurate results. Intuitively, comparing pixel

[2]https://github.com/subright/vision_search

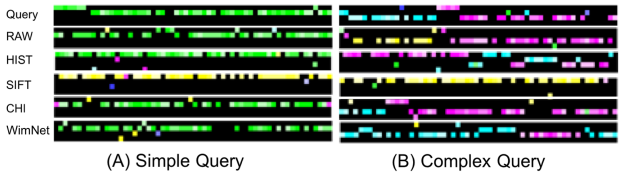

Figure 3: Top-1 result

by pixel cannot capture broad patterns in any dimensions. In the case of HIST, it is likely that the results are relatively accurate. However, it cannot capture specific pattern between devices, because the color histogram does not consider any dimensions. When we add more noises to web logs (+ 15% noise)[3], CHI cannot handle the noises to accurately find similar user behavior and generally WimNet produces more robust results overall.

4. CONCLUSIONS

In this work, we develop a convolution neural network, WimNet, which learns a latent representation from web log images. It is designed to capture user behavior across multiple devices and websites. To do this, we represent the web logs as an image by considering device and time dimensions as well as visiting sites with different color and brightness. According to the evaluation results with four baselines, WimNet produces the most robust and accurate top-k result regardless of a varying amount of noise. A main challenge of this work is that there is no ground-truth of similar users based on user behavior. Therefore, as a next step, we will evaluate our network on real-world web logs via both of qualitative and quantitative experiments. More specifically, we can conduct a user study to verify the effectiveness of our representation, or statistical test by computing correlation between business metrics (e.g., conversion rate and retention) and labels obtained based on our representation. Future work also includes extending and evaluating the performance of vision similarity search in other types of complex visualizations (e.g., treemaps, graph diagrams, and choropleths).

5. REFERENCES

[1] P. N. Bennett, R. W. White, W. Chu, S. T. Dumais, P. Bailey, F. Borisyuk, and X. Cui. Modeling the impact of short- and long-term behavior on search personalization. SIGIR '12, pages 185–194, 2012.

[2] F. Du, C. Plaisant, N. Spring, and B. Shneiderman. Finding similar people to guide life choices: Challenge, design, and evaluation. CHI '17, New York, NY, USA, 2017. ACM.

[3] A. Hassan, R. Jones, and K. L. Klinkner. Beyond dcg: User behavior as a predictor of a successful search. WSDM '10, pages 221–230, 2010.

[4] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, Nov. 2004.

[5] G. D. Montanez, R. W. White, and X. Huang. Cross-device search. CIKM '14, pages 1669–1678, 2014.

[6] G. Wang, X. Zhang, S. Tang, H. Zheng, and B. Y. Zhao. Unsupervised clickstream clustering for user behavior analysis. CHI '16, pages 225–236, 2016.

[3]These results can be found at the link above.