Identifying User Sessions in Interactions with Intelligent Digital Assistants

Rishabh Mehrotra*
Dept. of Computer Science
University College London
London, UK
r.mehrotra@cs.ucl.ac.uk

Ahmed El Kholy, Imed Zitouni,
Milad Shokouhi, Ahmed Hassan
Microsoft
Redmond, WA, USA
{ahkhol,izitouni,milads,hassanam}@microsoft.com

ABSTRACT

Search sessions have traditionally been considered as the focal unit of analysis for seeking behavioral insights from user interactions. While most session identification techniques have focused on the traditional web search setting; in this work, we instead consider user interactions with digital assistants (e.g. Cortana, Siri) and aim at identifying session boundary cut-offs. To our knowledge, this is one of the first studies investigating user interactions with a desktop based digital assistant. Historically, most user session identification strategies based on inactivity thresholds are either inherently arbitrary, or set at about 30 minutes. We postulate that such 30 minute thresholds may not be optimal for segregating user interactions with intelligent assistants into sessions. Instead, we model user-activity times as a Gaussian mixture model and look for evidence of a valley to identify optimal inter-activity thresholds for identifying sessions. Our results suggest a smaller threshold($\sim$2 minutes) for session boundary cut-off in digital assistants than the traditionally used 30 minute threshold for web search engines.

Keywords

digital assistants; user sessions; mixture models

1. INTRODUCTION

Intelligent digital assistants on mobile devices and computers, such as Cortana and Siri, have recently gained considerable attention as increasing number of users interact with them to fulfill their information needs and complete their tasks. Unlike traditional search engines, desktop based digital assistants provide plethora of avenues for users to verbally and textually interact with the system to perform a number of tasks like searching files, setting reminders, surfing the web or even chatting with the assistant. While these novel applications are useful and attractive for users, it is challenging for system designers to understand user interactions with such new interfaces and develop evaluation metrics.

Session identification is a common strategy used to develop metrics for web analytics and perform behavioral analyses of user-facing systems. Sessions allow us to look beyond individual queries, preserve semantic associations between query trails and maintain context of user activity. Strategies for session identification from log data have been extensively studied. Content based heuristics [6] exploit lexical content of queries for determining topical shift in query streams. Navigation-oriented heuristics [7] involve inferring browsing patterns based on the HTTP referrers and URLs associated with each request by a user. Time-oriented heuristics [2] refer to the assignment of an inactivity threshold between logged activities to serve as a session delimiter. Catledge & Pitkow [1] were among the first to use client-side tracking to examine browsing behavior and propose time based threshold. They reported that the mean time between logged events 9.3 minutes and chose to add 1.5 standard deviations to that mean to achieve a 25.5 minutes inactivity threshold. Over time this threshold has simplified to 30 minutes. This is the most popularly-used approach to identify sessions, with 30 minutes serving as the most common threshold [3, 7, 5].

Despite this interest in understanding the nature and manifestation of user sessions, no clear consensus about how to perform session identification has emerged. Further, most prior work has focused on user activity on search engines, insights from which may not be applicable to more novel interfaces like digital assistants. Intelligent assistants allow for radically new ways of information access, very different from traditional web search. As a result, it becomes important to explicitly consider such assistants and investigate user interactions with them to identify session boundaries.

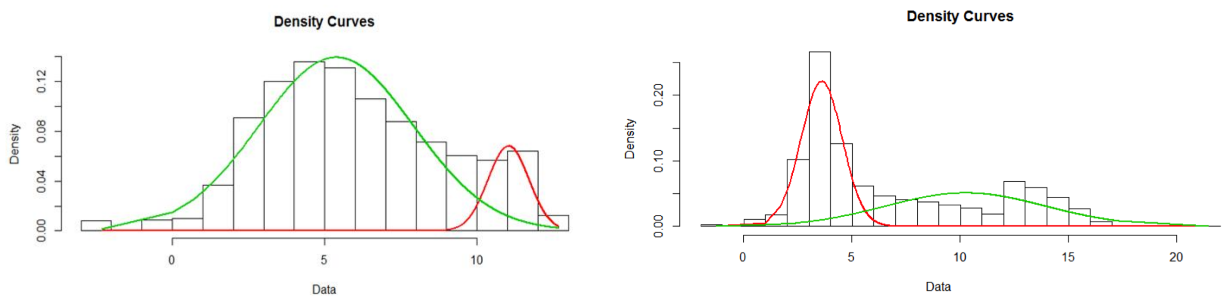
In this work, we consider a simple yet robust technique to identify sessions from user interactions with a desktop based digital assistant. We consider an approach based on Gaussian mixture models to model logarithmically scaled per user inter-activity times to obtain optimal inactivity thresholds to identify session boundaries. We base our experiments on real world log data and observe that the mixture model fits well to the Cortana desktop interaction data and that empirical data suggests a smaller time threshold ($\sim$2 minutes) for session cut-off.

2. IDENTIFYING SESSION BOUNDARY CUTOFF

We consider the temporal patterns in user initiated interactions and adopt an approach based on Gaussian mixture models to identify session boundary cutoff. We analyze a random sample of two weeks of Windows 10 Cortana app logs from June 2016, which contained user interaction data from over 3.6 million users, 21 million *impressions* and 6 million distinct queries. Each impression referred to one user initiating an interaction with Cortana and was tagged with a user ID and timestamp. We also use a random sample of Bing search logs from the same duration for comparison.

*Work conducted while at Microsoft Research.

WWW'17 Companion, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3054254



(a) Bing

(b) Cortana Desktop

Figure 1: Gaussian mixture model with bimodal components for Bing search engine and Cortana desktop. The two components correspond to (i) within session interactivity times and (ii) across session interactivity times.

Fitting Mixture of Gaussians:
We pre-process the data to obtain inter-activity times on a per user basis for all users, which helps us analyze the timegaps between the different user initiated interactions. We adopt an approach similar to Halfaker *et al.* [4] and plot a histogram based on the logarithmically scaled interactivity time and look for evidence of a valley. We fit a Mixture of Gaussians on logarithmically scaled inter-query times via Expectation-Maximization:

$$f(x,\theta) = \Sigma_{k=1}^{K} p_k \mathcal{N}(x; m_k \sigma_k) \tag{1}$$

where $\mathcal{N}(x; m_k\sigma_k)$ is a normal distribution with mean m_k and standard deviation σ_k and $K = 2$ for a 2-component mixture model. We follow the Expectation Maximization approach for parameter estimation wherein the goal is to maximize the likelihood function with respect to the parameters comprising the means and covariances of the components and the mixing coefficients.

E Step: for given parameter values we can compute the expected values of the posterior probabilities

$$p^i(k|n) = \frac{p_k^i \mathcal{N}(x; m_k\sigma_k)}{\Sigma_{k=1}^{K} p_k^i \mathcal{N}(x; m_k\sigma_k)} \tag{2}$$

M Step: re-estimate the parameters using the current posterior probabilities

$$m_k^{i+1} = \frac{\Sigma_{n=1}^{N} p^i(k|n) x_n}{\Sigma_{n=1}^{N} p^i(k|n)} \tag{3}$$

$$\sigma_k^{i+1} = \sqrt{\frac{1}{D} \frac{\Sigma_{n=1}^{N} p^i(k|n) \|x_n - m_k^{i+1}\|^2}{\Sigma_{n=1}^{N} p^i(k|n)}} \tag{4}$$

We identify an optimal inter-activity threshold for identifying sessions by finding the point where inter-activity time is equally likely to be within the first gaussians fit (within-session) and the second gaussians fit (between-session). The two mixture components fitted correspond to the within-session interactivity times and the between-session inter-activity times.

3. ANALYSIS

In Figure 1, we analyze the fitted Gaussian mixture model components fitted to the inter-activity times by plotting a log inter-activity time histogram overlaid with expectation maximization fits of a mixture of two log-normal cluster components. In order to better evaluate our approach, we first confirm known insights about user interaction with traditional search engines. Figure 1a presents the mixture model plot for a subset of Bing interaction data. The two mixture components intersect around 10.5 (log-scale), which translates to ∼24.1 minutes. This confirms known insights about session cut-off thresholds traditionally used in search engines. We perform the same analysis for Cortana desktop interactions and plot the mixture components in Figure 1b. From the intersection point of the two distributions (within session & between sessions), we observe a potential location for an inactivity cutoff and that a much smaller session boundary cut-off (∼ 2 minutes) is optimal for desktop based digital assistants. Indeed, most users use digital assistants to set reminders, alarms, and other small tasks which do not require sustained user interactions whereas web search tasks are often more complex, involving query reformulations and surveying.

4. CONCLUSION

We reconsidered the session identification thresholds for user interactions with digital assistants and empirically observed a much smaller threshold(∼2 minutes) for detecting session boundaries. As digital assistants become better at assisting users accomplish more complex tasks, the nature of user interactions and the corresponding sessions characteristics would change. As a result, the session cut-off thresholds would need to be revalidated using the same approach as described in this work. In future, we envision development of more sophisticated techniques for identification and evaluation of such session identification techniques, including task-dependent cut-offs and explicit evaluation, respectively.

5. REFERENCES

[1] L. D. Catledge and J. E. Pitkow. Characterizing browsing strategies in the world-wide web. *Computer Networks and ISDN systems*, 1995.

[2] R. Cooley, B. Mobasher, and J. Srivastava. Data preparation for mining world wide web browsing patterns. *Knowledge and information systems*.

[3] C. Eickhoff, J. Teevan, R. White, and S. Dumais. Lessons from the journey: a query log analysis of within-session learning. In *Proceedings of ACM WSDM*, 2014.

[4] A. Halfaker, O. Keyes, D. Kluver, J. Thebault-Spieker, T. Nguyen, K. Shores, A. Uduwage, and M. Warncke-Wang. User session identification based on strong regularities in inter-activity time. In *Proceedings of WWW*, 2015.

[5] J. L. Ortega and I. Aguillo. Differences between web sessions according to the origin of their visits. *Journal of Informetrics*, 2010.

[6] D. Shen, J.-T. Sun, Q. Yang, and Z. Chen. Building bridges for web query classification. In *Proceedings of ACM SIGIR*, 2006.

[7] M. Spiliopoulou, B. Mobasher, B. Berendt, and M. Nakagawa. A framework for the evaluation of session reconstruction heuristics in web-usage analysis. *Informs journal on computing*, 2003.