LUDWIG-MAXIMILIANS-UNIVERSITÄT AT MÜNCHEN
Department “Institut für Informatik”
Lehr- und Forschungseinheit Medieninformatik
Prof. Dr. Heinrich Hußmann

Masterarbeit

Automated UI Discovering for Web Application

Changkun Ou
hi@changkun.us

Bearbeitungszeitraum:	1.7.2018 bis 31.1.2019
Betreuer:	Malin Eiband and Dr. Daniel Buschek
Verantw. Hochschullehrer:	Prof. Dr. Heinrich Hußmann

Aufgabenstellung

Problem Statement Under standing user behavior helps designer optimize their product user experiences. At the same time, user benefits more productive from it. Since user intents are elusive, changeable and sometimes even undetermined, predict their behavior usually difficult and impoissible. In most cases, a user may performs a series of wasted actions before reach an intent destination. Nevertheless, user intents becomes clear step by step after performs a series of actions in a given context.

Scope of the Thesis To tackle the aforementioned challenges, the objective of this thesis is to develop a system that tracking user actions within a website, As a first step, literature review ... Based on the literature review, a intent model should be developed ... Then a system should implements the ... Nevertheless, user intents becomes clear step by step after performs a series of actions in a given context.

Tasks conduct a literature review to identify (research) question that simulates user input and predicts user ...

Requirements asdfasdf

Keywords asdfasdf

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbstständig angefertigt, alle Zitate als solche kenntlich gemacht sowie alle benutzten Quellen und Hilfsmittel angegeben habe.

München, December 12, 2018 . . . . . .

Acknowledgments

Thansk to everyone.

Abstract

This is a abstract.

Contents

1 Introduction

Introduction section [**?**]

1.1 Origin of Clickstream Research

1.2 This paper

2 Related works

Related works section

2.1 Client-side Clickstream

2.2 Productivity Quantification

2.3 Sequence to Sequence Learning

3 Experiment Design

TODO

3.1 Lab Study Design

3.2 Field Study

3.3 Dataset

4 Clickstream Models

4.1 URL2Vec Embedding

4.2 LSTM based Recurrent Network Architecture

5 Implementation

5.1 Client-side architecture

5.1.1 Browser Market Share

5.1.2 Architecture: Chrome as Example

5.2 Server-side architecture

5.2.1 Model Evolution Automation Architecture

5.3 Communication Protocol

6 Evaluation

6.1 Metrics

6.2 Model Evaluation

6.2.1 Prediction Accuracy

6.2.2 F1

6.2.3 t-SNE

6.3 Rationality of Designed Tasks

6.4 Explored Model Architecture Comparasion

6.5 Discussion

7 Applications

7.1 Client Side Browser Plugin

7.2 Standard Browser Web APIs

8 Conclusions

8.1 Summary

8.2 Future Works

Appendix

All resources relates to the thesis are open source, they can be found publicly in:

- Thesis homepage: `https://changkun.us/master-thesis-hci/`;
- GitHub repostory: `https://github.com/changkun/MasterThesisHCI/`.

A Content of enclosed USB

1. */documents/* - TODO

Bibliography

References

[1] Daniel Amo Filvá, Marc Alier Forment, Francisco Javier García Peñalvo, David Fonseca Escudero, and María José Casany Guerrero. Learning analytics to assess students behavior with scratch through clickstream. In *Proceedings of the Learning Analytics Summer Institute Spain 2018: León, Spain, June 18-19, 2018*, pages 74–82. CEUR-WS. org, 2018.

[2] Annika Baumann, Johannes Haupt, Fabian Gebert, and Stefan Lessmann. The price of privacy: An evaluation of the economic value of collecting clickstream data. *Business and Information Systems Engineering*, 02 2018.

[3] Fabrício Benevenuto, Tiago Rodrigues, Meeyoung Cha, and Virgílio Almeida. Characterizing user behavior in online social networks. In *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement*, IMC '09, pages 49–62, New York, NY, USA, 2009. ACM.

[4] Brodwin, D., D. O'Connell, and M. Valdmanis. Mining the Clickstream. pages 101–106, February 1995.

[5] Randolph E Bucklin and Catarina Sismeiro. How sticky is your web site? modeling site navigation choices using clickstream data. Technical report, Working paper, Anderson School UCLA, 2000.

[6] Nicholas G Carr. Hypermediation: commerce as clickstream. *Harvard Business Review*, 78(1):46–47, 2000.

[7] Ann Cavoukian. Should the oecd guidelines apply to personal data online. In *A report to the 22nd international conference of data protection commissioners*, 2000.

[8] Chatterjee, Patrali and Hoffman, Donna L and Novak, Thomas P. Modeling the clickstream: Implications for web-based advertising efforts. *Marketing Science*, 22(4):520–541, 2003.

[1] `http://creativecommons.org/licenses/by-nc-sa/4.0/`

[2] `https://github.com/changkun/MasterThesisHCI/blob/master/LICENSE`

[9] Yu Chi, Tingting Jiang, Daqing He, and Rui Meng. Towards an integrated clickstream data analysis framework for understanding web users' information behavior. *iConference 2017 Proceedings*, 2017.

[10] Richard J Courtheoux. Database marketing connects to the internet. *Interactive Marketing*, 2(2):129–137, 2000.

[11] Friedman, Wayne and Weaver, Jane. Calculating cyberspace: tracking "clickstreams.". February 1995.

[12] Susan E Gindin. Lost and found in cyberspace: Informational privacy in the age of the internet. *San Diego L. Rev.*, 34:1153, 1997.

[13] Avi Goldfarb. Analyzing website choice using clickstream data. In *The Economics of the Internet and E-commerce*, pages 209–230. Emerald Group Publishing Limited, 2002.

[14] Laxmi Amulya Gundala and Francesca Spezzano. Readers' demanded hyperlink prediction in wikipedia. In *Companion Proceedings of the The Web Conference 2018*, WWW '18, pages 1805–1807, Republic and Canton of Geneva, Switzerland, 2018. International World Wide Web Conferences Steering Committee.

[15] Nils Kammenhuber, Julia Luxenburger, Anja Feldmann, and Gerhard Weikum. Web search clickstreams. In *Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement*, IMC '06, pages 245–250, New York, NY, USA, 2006. ACM.

[16] Jerry Kang. Information privacy in cyberspace transactions. *Stan. L. Rev.*, 50:1193, 1997.

[17] Mingfeng Lin, Mei Lin, and Robert J. Kauffman. From clickstreams to searchstreams: Search network graph evidence from a b2b e-market. In *Proceedings of the 14th Annual International Conference on Electronic Commerce*, ICEC '12, pages 274–275, New York, NY, USA, 2012. ACM.

[18] Anália G. Lourenço and Orlando O. Belo. Catching web crawlers in the act. In *Proceedings of the 6th International Conference on Web Engineering*, ICWE '06, pages 265–272, New York, NY, USA, 2006. ACM.

[19] Joe Mandese. Clickstreams' in cyberspace. *Advertising Age*, 66(12):18–18, 1995.

[20] Florian Meier and David Elsweiler. Going back in time: An investigation of social media refinding. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '16, pages 355–364, New York, NY, USA, 2016. ACM.

[21] Bamshad Mobasher, Honghua Dai, Tao Luo, and Miki Nakagawa. Effective personalization based on association rule discovery from web usage data. In *Proceedings of the 3rd International Workshop on Web Information and Data Management*, WIDM '01, pages 9–15, New York, NY, USA, 2001. ACM.

[22] Chandramohan T N and Balaraman Ravindran. A neural attention based approach for clickstream mining. In *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data*, CoDS-COMAD '18, pages 118–127, New York, NY, USA, 2018. ACM.

[23] Novick, Bob. Internet Marketing: The Clickstream. `http:/www.im.com/archives/9503/0375.html`http:/www.im.com/archives/9503/0375.html, March 1995. Accessed: 2018-12-10.

[24] Balaji Padmanabhan, Zhiqiang Zheng, and Steven O. Kimbrough. Personalization from incomplete data: What you don't know can hurt. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '01, pages 154–163, New York, NY, USA, 2001. ACM.

[25] Jihyun Park, Kameryn Denaro, Fernando Rodriguez, Padhraic Smyth, and Mark Warschauer. Detecting changes in student behavior from clickstream data. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, LAK '17, pages 21–30, New York, NY, USA, 2017. ACM.

[26] Joseph Reagle and Lorrie Faith Cranor. The platform for privacy preferences. *Communications of the ACM*, 42(2):48–55, 1999.

[27] Joel R Reidenberg. Governing networks and rule-making in cyberspace. *Emory LJ*, 45:911, 1996.

[28] Joel R Reidenberg. Resolving conflicting international data privacy rules in cyberspace. *Stan. L. Rev.*, 52:1315, 1999.

[29] Narayanan Sadagopan and Jie Li. Characterizing typical and atypical user sessions in clickstreams. In *Proceedings of the 17th International Conference on World Wide Web*, WWW '08, pages 885–894, New York, NY, USA, 2008. ACM.

[30] Fabian Schneider, Anja Feldmann, Balachander Krishnamurthy, and Walter Willinger. Understanding online social network usage from a network perspective. In *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement*, IMC '09, pages 35–48, New York, NY, USA, 2009. ACM.

[31] Edith Schonberg, Thomas Cofino, Robert Hoch, Mark Podlaseck, and Susan L Spraragen. Measuring success. *Communications of the ACM*, 43(8):53–57, 2000.

[32] Atsushi Shimada, Yuta Taniguchi, Fumiya Okubo, Shin'ichi Konomi, and Hiroaki Ogata. Online change detection for monitoring individual student behavior via clickstream data on e-book system. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, LAK '18, pages 446–450, New York, NY, USA, 2018. ACM.

[33] Gavin Skok. Establishing a legitimate expectation of privacy in clickstream data. *Mich. Telecomm. & Tech. L. Rev.*, 6:61, 1999.

[34] Yanchun Sun and Chao Xin. Using coursera clickstream data to improve online education for software engineering. In *Proceedings of the ACM Turing 50th Celebration Conference - China*, ACM TUR-C '17, pages 16:1–16:6, New York, NY, USA, 2017. ACM.

[35] I-Hsien Ting, Chris Kimble, and Daniel Kudenko. Ubb mining: Finding unexpected browsing behaviour in clickstream data to improve a web site's design. In *Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence*, WI '05, pages 179–185, Washington, DC, USA, 2005. IEEE Computer Society.

[36] Luca Vassio, Idilio Drago, Marco Mellia, Zied Ben Houidi, and Mohamed Lamine Lamali. You, the web, and your device: Longitudinal characterization of browsing habits. *ACM Trans. Web*, 12(4):24:1–24:30, September 2018.

[37] Walsh, John and Godfrey, Sue. The Internet: a new era in customer service. *European Management Journal*, 18(1):85–92, 2000.

[38] Gang Wang, Xinyi Zhang, Shiliang Tang, Christo Wilson, Haitao Zheng, and Ben Y. Zhao. Clickstream User Behavior Models. *ACM Trans. Web*, 11(4):21:1–21:37, July 2017.

[39] Gang Wang, Xinyi Zhang, Shiliang Tang, Haitao Zheng, and Ben Y. Zhao. Unsupervised clickstream clustering for user behavior analysis. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 225–236, New York, NY, USA, 2016. ACM.

[40] Sarah Waterson, James A. Landay, and Tara Matthews. In the lab and out in the wild: Remote web usability testing for mobile devices. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '02, pages 796–797, New York, NY, USA, 2002. ACM.

[41] Sarah J. Waterson, Jason I. Hong, Tim Sohn, James A. Landay, Jeffrey Heer, and Tara Matthews. What did they do? understanding clickstreams with the webquilt visualization system. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '02, pages 94–102, New York, NY, USA, 2002. ACM.

[42] Tobias Weller. Compromised account detection based on clickstream data. In *Companion Proceedings of the The Web Conference 2018*, WWW '18, pages 819–823, Republic and Canton of Geneva, Switzerland, 2018. International World Wide Web Conferences Steering Committee.

[43] Toshihiko Yamakami. Inter-service revisit analysis of three user groups using intra-day behavior in the mobile clickstream. In *Proceedings of the 2009 International Conference on Hybrid Information Technology*, ICHIT '09, pages 340–344, New York, NY, USA, 2009. ACM.

[44] Zhi Yang, Christo Wilson, Xiao Wang, Tingting Gao, Ben Y. Zhao, and Yafei Dai. Uncovering social network sybils in the wild. *ACM Trans. Knowl. Discov. Data*, 8(1):2:1–2:29, February 2014.

[45] Jan Zaloudek. User Behavior Clustering and Behavior Modeling Based on Clickstream Data. Master's thesis, Czech Technical University in Prague, Faculty of Electrical Engineering Department of Computer Science, May 2018.

[46] Xiang Zhang, Hans-Frederick Brown, and Anil Shankar. Data-driven personas: Constructing archetypal users with clickstreams and user telemetry. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 5350–5359, New York, NY, USA, 2016. ACM.