

Advanced Click Models and their Applications to IR

SIGIR 2015 Tutorial

Aleksandr Chuklin
Google Switzerland &
University of Amsterdam
Zürich, Switzerland
a.chuklin@uva.nl

Ilya Markov
ISLA, University of Amsterdam
Amsterdam, The Netherlands
i.markov@uva.nl

Maarten de Rijke
ISLA, University of Amsterdam
Amsterdam, The Netherlands
derijke@uva.nl

ABSTRACT

This tutorial concerns with more advanced and more recent topics in the area of click models. Here, we discuss recent developments in the area with a particular focus on applications of click models. The tutorial features a guest talk and a live demo where participants have a chance to build their own advanced click model.

While this is the second part of the two half-day tutorials, it is not required for participants to attend the first one. In the beginning of this part, a short introduction to basic click models will be given so that all participants share a common vocabulary. Then, recent advances in click models will be discussed.

Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.3 Information Search and Retrieval

Keywords

Click models; web search

1. INTRODUCTION

Click models, probabilistic models of the behavior of search engine users, have been studied extensively by the information retrieval community during the last five years.

A good body of work has been published on click models over the past 5 to 10 years. Importantly, the area continues to be an active one, with new models aimed at describing and/or predicting behavioral data being proposed at each of the main information retrieval conferences (SIGIR, WSDM, WWW, CIKM) over the past few years. There is also a fair number of publications that use click models presented in the main information retrieval conferences as well as ACM journals such as TOIS. All this indicates that there is a substantial and continued level of interest in the topic.

Indeed, we now have a handful of basic click models, inference methods, evaluation principles and applications for click models, that form the building blocks of ongoing research efforts in the area. The time is right to organize and present this material to a broad audience of interested information retrieval researchers, whether

junior or senior. Many of the click models available today have been proposed by different industrial and academic research groups—one of the key aims of our proposed tutorial is to bring these together and offer a unified perspective. To achieve this, we describe the basic click models, inference methods and evaluation principles. We supplement this with an account of available datasets and packages plus a live demo based on these. We also present click model applications accompanied by examples.

We expect the tutorial to be useful for both researchers and practitioners that either want to develop new click models, use them in their own research in other areas or apply the models described here to improve actual search systems.

2. OBJECTIVES

A large body of research on click models has been developed. This research has improved our understanding of user behavior in web search and facilitated the usage of click models in various search-related tasks. Current studies use a broad range of notations and terminology, perform experiments using different and mostly proprietary datasets, do not detail the model inference procedures used and, thus, do not provide a general systematic view on the research area. This, in turn, slows down the development and hinders the application of click models. The goal of this tutorial is to bring together current efforts in the area, summarize the research performed so far and give a holistic view on existing click models for web search. More specifically, the aims of this tutorial are the following:

1. Describe existing click models in a unified way, i.e., using common notation and terminology, so that different models can easily be related to each other.
2. Compare commonly used click models, discuss their advantages and limitations and provide a set of recommendations on their usage.
3. Provide ready-to-use formulas and implementations of existing click models and detail general inference procedures to facilitate the development of new ones.
4. Give an overview of existing datasets and tools for working with click models and develop new ones.
5. Provide an overview of click model applications and directions for future development of click models.

Our target audience consists of researchers and developers in information retrieval who are interested in formally capturing user interactions with search engine result pages, whether for ranking purposes, to simulate user behavior in a lab setting, or simply to gain

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author(s). Copyright is held by the owner/author(s).

SIGIR'15, August 09-13, 2015, Santiago, Chile.

ACM 978-1-4503-3621-5/15/08.

DOI: <http://dx.doi.org/10.1145/2766462.2767882>.

deeper insights in user behavior and interaction data. The tutorial will be useful as an overview for anyone starting research work in the area as well as for practitioners seeking concrete recipes.

The tutorial aims to provide a map of an increasingly rich landscape of click models. By the end of the half-day tutorial, attendees should be familiar with the basic definitions and intuitions of what we consider to be the core models, with inference tasks for these models, and with uses of these models. While our presentation is necessarily formal in places, we make a serious effort to relate the models, the inference procedures and the applications back to the core information retrieval task and to web search data by including a fair number of examples. We hope that this supplies attendees who are new to the area with effective means to start using click models in their own research.

3. DETAILED SCHEDULE

The advanced tutorial consists of two blocks of 1.5 hours. Each block will incorporate three types of activity: (1) presentation of the material, (2) live coding sessions, and (3) discussions at the end of each 1.5 hours. Additionally we plan to bring an author of a recent click model paper to give a short presentation and participate in the discussion. For the coding sessions, we will provide code examples and data samples. The participants will be able to either follow the examples on the slides or perform them live along with the presentation. During the discussion sessions at the end of each 1.5 hour block, we will offer the participants a number of quizzes aimed to (self-)assess the comprehension of the presented material. The schedule of the advanced tutorial is as follows.

Block 1 (1.5 hours)

1. Overview of Advanced Click Models

Here we present main development directions of click models and discuss most prominent members of each class. We discuss models' assumptions, observed and hidden random variables, models' parameters and relations between parameters and random variables. This discussion is concluded with a summary of differences and similarities between the presented models. The following types of click models will be covered in this part:

- (a) Click models for aggregated search
- (b) Beyond a single query session
- (c) User diversity
- (d) Eye tracking and mouse movements
- (e) Nonlinear search engine result page examination
- (f) Using editorial judgements

2. Live Demo: Building an Advanced Click Model

In this part, we use samples of publicly available datasets and open-source software packages to guide the participants through a process of extending a basic click model to incorporate some new ideas. In particular, we will use an open source package maintained by Markov¹ as well as samples from interaction data shared by Yandex.²

3. Recap and Discussion

Coffee Break (30 minutes)

Block 2 (1.5 hours)

4. Live Demo: Building an Advanced Click Model (continued)
5. Guest talk: "Building a Click Model: From Idea to Implementation"
6. Click Model Applications and Future Directions In this part we discuss different applications for modern click models in different areas such as ranking, evaluation or user simulation. We also discuss future directions for click models, the area we believe to stay active in the coming years with more ideas coming in and being integrated and re-integrated.
 - (a) Applications
 - i. Inference of Document Relevance
 - ii. Click Model-based Metrics
 - iii. Simulating Users
 - (b) Future Directions
 - i. Rich User Feedback
 - ii. Heterogeneous Search
 - iii. Cross-Device and Cross-Session Search Tasks
7. Recap and Discussion

4. TYPE OF SUPPORT MATERIALS TO BE SUPPLIED TO ATTENDEES

- Complete draft of the book [1]
- Copy of the slides
- Code and data samples to follow live demos

Acknowledgements

This work was partially supported by the grant P2T1P2_152269 of the Swiss National Science Foundation, the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement nr 312827 (VOX-Pol), the Netherlands Organisation for Scientific Research (NWO) under project nrs 727.011.005, 612.001.116, HOR-11-10, 640.006.013, 612.066.930, CI-14-25, SH-322-15, Amsterdam Data Science, the Dutch national program COMMIT, the ESF Research Network Program ELIAS, the Elite Network Shifts project funded by the Royal Dutch Academy of Sciences (KNAW), the Netherlands eScience Center under project number 027.012.105, the Yahoo! Faculty Research and Engagement Program, the Microsoft Research PhD program, and the HPC Fund.

All content represents the opinion of the authors which is not necessarily shared or endorsed by their respective employers and/or sponsors.

REFERENCES

- [1] A. Chuklin, I. Markov, and M. de Rijke. *Click Models for Web Search*. Morgan & Claypool, 2015. To appear.

¹<https://github.com/markovi/PyClick>

²<http://imat-relpred.yandex.ru/en/datasets>