Click models for web search and their applications to IR

WSDM 2016 Tutorial

Chuklin, A.; Markov, I.; de Rijke, M.

Published in:
WSDM'16

DOI:
10.1145/2835776.2855113

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
We show how the framework of probabilistic graphical models
well as ACM journals such as TOIS. All this indicates that there is
The time is right to organize and present this material to a broad
ABSTRACT
In this tutorial we give an overview of click models for web search. We
show how the framework of probabilistic graphical models helps to explain user behavior, build new evaluation metrics and
perform simulations. The tutorial discusses foundational aspects alongside experimental details and applications, with live demos
and discussions of publicly available resources.

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 decade.
A good body of work has been published on click models over the past five to ten 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 prac-
titioners 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 hin-
ders 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 advan-
tages and limitations and provide a set of recommendations
on their usage.

3. Provide ready-to-use formulas and implementations of exist-
ing 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 direc-
tions for future development of click models.

Our target audience consists of researchers and developers in in-
formation 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
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 land-
scape 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 tutorial is organized in two blocks of 1.5 hrs.

**Block 1 (1.5 hours)**

1. Introduction, historical notes, aims.

2. Basic click models

   Here, we describe basic click models for web search, starting with the simplest random click model and arriving to the more sophisticated DCM, DBN and UBM models [1, 3, 4]. We discuss models’ assumptions, observed and hidden random variables, models’ parameters and relations between parameters and random variables. We also demonstrate how to infer click probabilities (either full or conditional) and simulate the users’ clicks.

3. Click model parameter estimation (training)

   We describe the process of training click model parameters from past click observations. We review two main parameter estimation techniques, namely maximum likelihood estimation (MLE) and the expectation-maximization algorithm (EM). We then give several examples of MLE and EM parameter estimation for existing click models to support the theory.

4. Click model evaluation

   We discuss different ways to evaluate click models and to compare them to each other. We start with traditional approaches, such as log-likelihood/perplexity evaluation and click-through rate prediction, and then discuss more application-oriented evaluation methods like NDCG evaluation.

**Block 2 (1.5 hours)**

5. Data and tools

   Here, we discuss publicly available datasets for experimenting with click models. We also describe open-source software packages and libraries that facilitate working with click models.

6. Live Demo

   During the live demo we compare the click models, discussed during the first block, using the presented evaluation methods. We also discuss the differences between parameter estimation techniques from the practical point of view (e.g., learned parameters, learning time). The demo uses an open source package maintained by Markoi as well as samples from interaction data shared by Yandex.

7. Advanced click models

   Here we present a brief overview of main development directions of click models and discuss most prominent members of each class. We discuss models’ assumptions and main applications.

8. Click model applications

   In this part, we discuss different applications for modern click models in different areas such as ranking, evaluation or user simulation. We augment the theoretical discussion with the practical demonstrations of some of the applications.

9. Recap and discussion

4. TYPE OF SUPPORT MATERIALS TO BE SUPPLIED TO ATTENDEES

   - The authors’ version of the book on click models [2]
   - Copy of the slides
   - Code and data samples to follow live demos and for offline experimentation

We also maintain a web site with all resources related to this tutorial at [http://clickmodels.weebly.com](http://clickmodels.weebly.com).

Acknowledgements

The tutorial is partially based on research that was supported by Amsterdam Data Science, the Dutch national program COMMIT, Elsevier, the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement nr 312827 (VOX-Pol), the ESF Research Network Program ELIAS, the Royal Dutch Academy of Sciences (KNAW) under the Elite Network Shifts project, the Microsoft Research Ph.D. program, the Netherlands eScience Center under project number 027.012.105, the Netherlands Institute for Sound and Vision, 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, 652.002.-001, the Yahoo Faculty Research and Engagement Program, and Yandex. All content represents the opinion of the authors, which is not necessarily shared or endorsed by their respective employers and/or sponsors.

REFERENCES


---

1 [https://github.com/markovi/PyClick](https://github.com/markovi/PyClick)