Dynamic Information Retrieval Modeling

Hui Yang
Georgetown University
huiyang@cs.georgetown.edu

Marc Sloan
University College London
m.sloan@cs.ucl.ac.uk

Jun Wang
University College London
junwang@cs.ucl.ac.uk

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Retrieval models; Relevance feedback; Search process

Keywords

Dynamic Information Retrieval Modeling; Probabilistic Relevance Model; Reinforcement Learning

ABSTRACT

In Dynamic Information Retrieval modeling we model dynamic systems which change or adapt over time or a sequence of events using a range of techniques from artificial intelligence and reinforcement learning. Many of the open problems in current IR research can be described as dynamic systems, for instance, session search or computational advertising. State of the art research provides solutions to these problems that are responsive to a changing environment, learn from past interactions and predict future utility. Advances in IR interface, personalization and ad display demand models that can react to users in real time and in an intelligent, contextual way.

The objective of this half-day tutorial is to provide a comprehensive and up-to-date introduction to Dynamic Information Retrieval Modeling. We motivate a conceptual model linking static, interactive and dynamic retrieval and use this to define *dynamics* within the context of IR. We then cover a number of algorithms and techniques from the artificial intelligence (AI) and online learning literature such as Markov Decision Processes (MDP) [1], their partially observable variation (POMDP) [5] and multi-armed bandits [7].

Following this we describe how to identify dynamics in an IR problem and demonstrate how to model them using the described techniques. The remainder of the tutorial will then cover an array of state-of-the-art research on dynamic systems in IR and how they can be modeled using using dynamic IR [2, 6]. We use research on session search [3], multi-page search [4] and online advertising [8] as in-depth examples of such work.

This tutorial is of relevance to IR practitioners and researchers, where we will present the merits of dynamic information retrieval modeling and introduce the relevant techniques. The content will be of particular interest to researchers working in the areas of statistical modeling, personalization and recommendation, and is also relevant to practitioners in Web search, online advertising and anyone who works with big data.

After this tutorial, attendees will:

- Be able to identify the dynamics in an IR system
- Be able to model these dynamics using techniques from AI and reinforcement learning
- Have knowledge of the state-of-the-art research in dynamic information retrieval modeling

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 author/owner(s).
WSDM'15, February 2–6, 2015, Shanghai, China.
ACM 978-1-4503-3317-7/15/02.
http://dx.doi.org/10.1145/2684822.2697038.

1. SPEAKERS

Grace Hui Yang is an Assistant Professor in the Department of Computer Science at Georgetown University. Grace obtained her PhD from School of Computer Science, Carnegie Mellon University. Grace's current research interests include session search, search engine evaluation, privacy-preserving information retrieval, and information organization. Prior to this, she conducted research on question answering, ontology construction, near-duplicate detection, multimedia information retrieval and opinion and sentiment detection. The results of her research have been published in SIGIR, CIKM, ACL, TREC, ECIR, and WWW since 2002. Grace serves on the Editorial Board for Information Retrieval Journal. She also served as an area chair/senior program committee member for SIGIR 2014 and co-chaired the SIGIR 2013 and SIGIR 2014 Doctoral Consortium. She is a program committee member in SIGIR, ACL, EMNLP, CIKM, WSDM, and KDD. Home page: `http://www.cs.georgetown.edu/~huiyang`

Marc Sloan is a senior graduate student completing his PhD thesis in University College London. He has interned at Microsoft Research working on contextual, session based search result blending. His background is in information retrieval, machine learning and financial computing and his thesis is on dynamic IR. His research interests in information retrieval include applying reinforcement learning techniques such as multi-armed bandits and POMDPs to IR learning systems over time, contextual session search and query suggestion. Marc has published and presented reinforcement learning in IR research at top-tier conferences and workshops such as SIGIR, WWW, WSCD and the Large-scale On-

line Learning and Decision Making Workshop. Homepage: http://mediafutures.cs.ucl.ac.uk/people/MarcSloan/

Jun Wang is Senior Lecturer (Associate Professor) in Computer Science, University College London and the Founding Director of the MSc/MRes Web Science & Big Data Analytics programme. His main research interests are statistical modelling of information retrieval, collaborative filtering and computational advertising. He was a recipient of the SIGIR Doctoral Consortium Award in 2006, the Microsoft Beyond Search award in 2007 and won the Best Paper Prizes in ECIR09 and ECIR12. JunâĂŹs recent service includes (Senior) PCs for SIGIR, CIKM, WSDM, and RecSys, and presenting an ECIR2011 tutorial on Risk Management in Information Retrieval, a CIKM2011 tutorial on statistical IR modeling and a CIKM2013 tutorial on Computational Advertising. Home page: http://www.cs.ucl.ac.uk/staff/J.Wang/

2. GUEST SPEAKER

Maarten de Rijke is full professor of Information Processing and Internet in the Informatics Institute at the University of Amsterdam. De Rijke leads the Information and Language Processing Systems group, one of the world's leading academic research groups in information retrieval. During the most recent computer science research assessment exercise, the group achieved maximal scores on all dimensions. His research focus is on intelligent information access, with projects on self-learning search engines, semantic search, and social media analytics. A Pionier personal innovational research incentives grant laureate (comparable to an advanced ERC grant), De Rijke has generated over 35MEuro in project funding. With an h-index of 50 he has published over 600 papers, published or edited over a dozen books, is the Editor-in-Chief of ACM Transactions on Information Systems and of Springer's Information Retrieval book series, (associate) editor for various journals and book series, and a current and former coordinator of retrieval evaluation tracks at TREC, CLEF and INEX. He was co-chair for SIGIR 2013, general chair for ECIR 2014 and will be program co-chair for information retrieval for CIKM 2015 and short paper co-chair for SIGIR 2015. He is the director of the University of Amsterdam's Intelligent Systems Lab (ISLA), of the Center for Creation, Content and Technology (CCCT), of the Ad de Jonge Center for Intelligence and Security Studies, and of the Amsterdam Data Science research center. The retrieval and language technology developed by his research group is being used by organizations around the Netherlands and beyond, and has given rise to various spin-off initiatives. Homepage: https://staff.fnwi.uva.nl/m.derijke/

3. REFERENCES

[1] Bellman, R. A markovian decision process. *Indiana University Mathematics Journal 6* (1957), 679–684.

[2] Fetterly, D., Manasse, M., Najork, M., and Wiener, J. A large-scale study of the evolution of web pages. In *WWW '03*, pp. 669–678.

[3] Guan, D., Zhang, S., and Yang, H. Utilizing query change for session search. In *SIGIR '13*, pp. 453–462.

[4] Jin, X., Sloan, M., and Wang, J. Interactive exploratory search for multi page search results. In *WWW '13*, pp. 655–666.

[5] Kaelbling, L. P., Littman, M. L., and Cassandra, A. R. Planning and acting in partially observable stochastic domains. *Artificial intelligence 101*, 1 (1998), 99–134.

[6] Leskovec, J., Backstrom, L., and Kleinberg, J. Meme-tracking and the dynamics of the news cycle. In *KDD '09*, pp. 497–506.

[7] Robbins, H. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society 58*, 5 (Sep 1952), 527–535.

[8] Yuan, S., and Wang, J. Sequential selection of correlated ads by pomdps. In *CIKM '12*, pp. 515–524.