Dynamic Information Retrieval Modeling

Hui Yang
Georgetown University, USA
huiyang@cs.georgetown.edu

Marc Sloan, Jun Wang
University College London, UK
{M.Sloan, J.Wang}@cs.ucl.ac.uk

ABSTRACT

Dynamic aspects of Information Retrieval (IR), including changes found in data, users and systems, are increasingly being utilized in search engines and information filtering systems. Existing IR techniques are limited in their ability to optimize over changes, learn with minimal computational footprint and be responsive and adaptive. The objective of this tutorial is to provide a comprehensive and up-to-date introduction to Dynamic Information Retrieval Modeling, the statistical modeling of IR systems that can adapt to change. It will cover techniques ranging from classic relevance feedback to the latest applications of partially observable Markov decision processes (POMDPs) and a handful of useful algorithms and tools for solving IR problems incorporating dynamics.

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

1. INTRODUCTION

Big data and human-computer information retrieval (HCIR) are changing IR: they capture the dynamic changes in the data and dynamic interactions of users with IR systems. A dynamic system is one which changes or adapts over time or a sequence of events. Many modern IR systems and data exhibit these characteristics which are largely ignored by conventional techniques. What is missing is an ability for the model to change over time and be responsive to stimulus. Documents, relevance, users and tasks all exhibit dynamic behavior that is captured in data sets typically collected over long time spans and models need to respond to these changes. Additionally, the size of modern datasets enforces limits on the amount of learning a system can achieve. Further to this, 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 tutorial is to provide a comprehensive and up-to-date introduction to Dynamic IR Modeling. We will define *dynamics*, what it means within the context of IR and highlight examples of problems where dynamics play an important role. Following this we will introduce a handful of theories and algorithms which can be used to incorporate dynamics into an IR model before presenting an array of state-of-the-art research that already does, such as in the areas of session search [4] and online advertising [6].

The theoretical component will be based around the Markov Decision Process (MDP) [3], a mathematical framework taken from the field of Artificial Intelligence (AI) that enables us to construct models that change according to sequential inputs. We will define the framework and the algorithms commonly used to optimize over it [2] and generalize it to the case where the inputs aren't reliable [5]. We will explore the topic of reinforcement learning more broadly and introduce another tool known as a Multi-Armed Bandit [1] which is useful for cases where exploring model parameters is beneficial.

After this tutorial, attendees will:

- Have a sound understanding of how to identify the dynamic aspects of an IR problem
- Be able to utilize a Markov Decision Process in their IR models so as to incorporate dynamic elements
- Have knowledge of how such models are already used in state-of-the-art IR research

SIGIR'14, July 6–11, 2014, Gold Coast, Queensland, Australia.
ACM 978-1-4503-2257-7/14/07.
http://dx.doi.org/10.1145/2600428.2602297.

2. REFERENCES

[1] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Mach. Learn.*, 47(2-3):235–256, May 2002.

[2] R. Bellman. *Dynamic Programming.* Princeton University Press, Princeton, NJ, USA, first edition, 1957.

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

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

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

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