

Building Cost Models of Information Interaction & Retrieval

Full Day Tutorial

Leif Azzopardi
University of Strathclyde
Glasgow, United Kingdom
leifos@acm.org

ABSTRACT

Modeling how people interact with search interfaces has been of particular interest and importance to the field of Interactive Information Retrieval. Recently, there has been a move to developing formal models of the interaction between the user and the system, whether it be to: (i) run a simulation, (ii) conduct an economic analysis, (iii) measure system performance, or (iv) simply to better understand the interactions and hypothesise about user behaviours. In such models, they consider the costs and the benefits that arise through the interaction with the interface/system and the information surfaced during the course of interaction. In this full day tutorial, we will focus on describing a series of cost (and benefit) models that have been proposed in the literature and how they have been applied in various scenarios. The tutorial will be structured into two parts. First, we will provide an overview of Decision Theory and Cost-Benefit Analysis techniques, and how they can and have been applied to a variety of Interactive Information Retrieval scenarios. For example, *how do we reduce the cost browsing facets? how many results should we show per page? and how do we get users to pose longer queries?* The second part of the day will be dedicated to building cost models where we will discuss different techniques to build and develop models from which one can draw testable hypotheses from. In this more practical session, we will also discuss how costs and benefits can be estimated, and how the models can help inform and guide experimentation. During the tutorial participants will be challenged to build cost models for a number of problems (or even bring their own problems to solve). We will then conclude with presentations of solutions followed by a summary and overview of challenges and future directions. This tutorial is aimed at participants wanting to know more about cost models and how they can be applied, interpreted and built in the context of information seeking, search and retrieval.

Keywords

Search Behaviour, User Models, Retrieval Strategies, Evaluation

1. OVERVIEW AND BACKGROUND

An emerging area of research within Interactive Information Retrieval (IIR) is the development of cost models that characterize the cost of interaction with a particular interface or particular features within the interface. Such cost models have been used for various purposes, such as: (i) controlling the number of interactions a simulated user can perform in a given period of time [13, 14], (ii) approximating the time spent examining the results in a ranked list, e.g., within time biased measures [31, 33], (iii) analyzing and empirically evaluating the costs and benefits of different interfaces [7, 24, 28] (iv) estimating the cost of different courses of interaction to determine the most efficient course of action [4, 23, 27]. Essentially, these lines of research have aimed at measuring the *cost* that the user incurs, as opposed to traditional performance measures, which only look at measuring the *gain* or *benefit* that the user receives from such actions. Thus such models/measures of cost complement and extend the existing research on measuring the value of searching from a user perspective [9]. IIR systems often provide various features and functions, such as query suggestions and relevance feedback, that a user may or may not decide to use. The decision to take such a particular option has associated costs and may lead to some benefit. Thus, a savvy user would take decisions that maximises their net benefit. Given this premise, it is possible to design interfaces that reduce costs and increase benefits. Further, also helps to contextualise and understand why users select one action over another and hypothesise about how a user would interact under particular conditions. Through the tutorial we will describe an array of models and draw a number of insights into different decisions, provide explanations for observed behaviours and show how it is possible to generate numerous testable hypotheses from the models. The tutorial will be composed of two parts: (1) Theory and (2) Practice. In the theory sessions we will cover how: cost-benefit analysis and decision theory, and how it can be applied, followed by descriptions and explanations of a series of cost (and benefit) models that have been developed in the context of interactive information retrieval and searching. In the practical sessions we will focus on explaining how to building such models, how to use the models in practice, along with a discussion on how to estimate and measure costs/benefits in IIR. Then participants, in groups, will undertake a series of activities, where

they are challenged to build and develop a cost model for IIR. *Participants are free to bring along their own problem, or work on the problems provided in class.* The tutorial will conclude with a discussion of solutions along with a summary of the challenges and future directions for modeling costs in IIR.

2. FORMAT AND SCHEDULE

As previously mentioned the first half of the day will be focused on theory and existing work, and the second half of the day will be focused on how to build such models and a practical session.

2.1 Introduction and the Economics of Search

In the first session we will briefly provide the context and motivation for developing formal models of Information Seeking and Retrieval (ISR). To begin, we will discuss and describe various conceptual and descriptive models of ISR, including Bates' Berry Picking Model [15] and the ISR framework proposed by Ingwersen and Kalvero [21], along with other such models (e.g. [37, 25, 18]). This will provide the background for the tutorial where we point out the limitations of such models and motivate the need to develop models that are not only descriptive in nature, but which help us to hypothesise and explain observed behaviours. We will then provide an overview of Cost-Benefit Analysis techniques and Decision Theory, before describing a series of different models which attempt to answer various questions like:

- who should search more the student or the professor? [17]
- how should facets be organised to reduce the cost of interaction? [23]
- how long a user's query should be? [10]
- should the user pose a specific or vague query?
- should the user take a suggestion or re-formulate?
- when should a user employ relevance feedback? [22]
- when would the "find similar" functionality be worthwhile to the user?
- when should a user stop searching? [16, 32]
- should a user examine many documents and issue few queries, or vice versa? [2, 4]

In describing the different models, we will explore and explain how different parameters affect the decisions at play. This will provide a number of insights into how these models can be used to inform experimentation, and guide further research.

2.2 How To Model and Practical Sessions:

In the next session (after lunch), we will focus on how to build cost models and how to use them to generate hypotheses using various methods (analytical, graphical and computational) [36, 26]. The essence of such models is to develop a cost function and a benefit function given the different interactions that users can perform. Given these functions, it is then possible to establish under what conditions cost is

minimized and/or benefit is maximized. Alternatively, given the cost/benefit function for various interaction/choices it is possible to compare when one choice is preferable to another. This section will explore the following topics:

- Decision Making and Cost Models [20, 26].
- Optimality and Rationality [30, 29, 34].
- What choices are available? What are the limitations?
- What are the interactions, costs and benefits?
- What will vary? Parameters?
- Define the problem and the goal
- Construct a cost and benefit function
- What method to choose: analytical, graphically, computational
- Solve, plot, compute
- Draw inferences and generating hypotheses
- How to estimate costs / benefits
- How to design/construct experiments based upon the models

To help illustrate how to undertake the model building process, we will use an example based on the simple scenario of finding the first highly relevant document in a result list [11]. We will describe how we can characterize the problem, enumerate the different variables, and show how the different variables impact the overall cost and the design of the interface. Given this scenario, we will show how we can devise a number of experiments to empirically test and explore the model - and discuss how it is possible to estimate various costs and benefits.

The final session of the tutorial will be dedicated to building models. For example, participants will be given a context such as a user trying to find an app on a mobile phone or tablet. The goal here is to find the app that the user wants to use. So the focus will be on building a cost function to model the different ways in which the user can find the app, e.g. search or browse. To add realism to the scenario, we will consider different types of users, ones that can remember where the app is, and those who cant (i.e., best and average/worse cases) as well as consider how screen size and app icon size can be modelled to arrive at an optimal size and layout for such interaction. During the practical session, participants will be encouraged to abstract away the details to form a representation of the problem, and identify the main variables that are likely to influence the interaction (i.e., the number of apps on the phone, the number of apps per screen page, the cost of moving between screens, etc.). These will be used to formulate the costs, then we will be able to reason about when it is better to search for an app and when it is better to browse. We will also consider alternative designs, such as presenting the most used apps first, or a hierarchical browsing structure, and whether they are likely to be more efficient or not, or under what circumstances.

3. INTENDED LEARNING OUTCOMES

By the end of the tutorial, participants should be able to:

- Understand the context and rationale for formal models;
- Describe and explain different types of cost/benefit models;
- Define and create a cost/benefit model;
- Explain and infer the predicted user behaviour given the cost models;
- Generate hypotheses regarding user behavior given cost/benefit models.

4. PRE-REQS, AUDIENCE, LENGTH, ETC

A basic understanding of Information Retrieval will be assumed. That is, we expect the attendees to know what an IR system is, the inputs and outputs of the system, the standard ways to evaluate a retrieval system, along with some knowledge of the different types of retrieval models (i.e., vector space, best match, etc.) [35]. Some basic math knowledge is helpful (i.e., differentiation, etc.), but not required as key concepts will be explained along with alternatives methods, i.e., computational methods and simulation.

The intended audience would be students undertaking a PhD, and researchers that are interested in how to formalize the problems that they are working on, specifically to inform their experimentation, regarding user behaviour and interaction. The tutorial is designed to be introductory progressing to intermediate.

The proposed tutorial is for a full day.

We will distribute slides and bibliography to participants along with exercise sheets for in-class exercises.

This tutorial focuses in on cost models, specifically, and thus is related to the series of tutorials on formal models, in general, that we have presented at the Information Foraging Summer School (2013-2014), a mini-tutorial on Information Foraging Theory was presented at IliX 2014, and full day tutorials at ACM SIGIR 2015, ACM CIKM 2015, and an advanced modelling course at ACM ICTIR 2016. In these previous full day tutorials, the focus was on providing a comprehensive overview of the different types of models, i.e. economic models, information foraging theory, interactive Probability Ranking Principle, and the Card Model. However this tutorial steps back and focuses on the fundamentals i.e. of modelling costs and benefits.

5. BIOGRAPHY

Leif Azzopardi has been recently awarded a Chancellor's Fellowship at the University of Strathclyde within the Department of Computer and Information Sciences. Prior to this appointment, he was a Senior Lecturer in the School of Computing Science at the University of Glasgow. His research focuses on building formal models for Information Retrieval - usually drawing upon different disciplines for inspiration, such as Quantum Mechanics, Operations Research, Microeconomics, Transportation Planning and Gamification. Central to his research is the theoretical development of statistical language models for Information Retrieval, where his research interests include:

- Models for the retrieval of documents, sentences, experts and other information objects [12, 19];
- Probabilistic models of user interaction and the simulation of users for evaluation [1, 5, 6];
- Microeconomic models of information interaction, specifically how cost and effort affect interaction and performance with search systems [2];
- Methods which assess the impact of search technology on society in application areas such as, search engine bias and the accessibility of e-Government information [8], and;
- Search for fun (i.e. the SINS of users) [3].

He received his Ph.D. in Computing Science from the University of Paisley in 2006, and he received a First Class Honours Degree in Information Science from the University of Newcastle, Australia, 2001. In 2010, he received a Post-Graduate Certificate in Academic Practice and has been lecturing at the University of Glasgow since then. He has given numerous invited talks on Formal Models of Information Seeking and Retrieval throughout the world and lectured at the Information Foraging Summer School (2011, 2012 and 2013) and Symposium of Future Directions in Information Access (2007-2013).

6. REFERENCES

- [1] L. Azzopardi. Query side evaluation: an empirical analysis of effectiveness and effort. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 556–563. ACM, 2009.
- [2] L. Azzopardi. The economics in interactive information retrieval. In *Proc. of the 34th international ACM SIGIR conference*, pages 15–24. ACM, 2011.
- [3] L. Azzopardi. Searching for unlawful carnal knowledge. In *Proceedings of the SIGIR Workshop: Search for Fun*, volume 11, pages 17–18, 2011.
- [4] L. Azzopardi. Modelling interaction with economic models of search. In *Proc. of the 37th ACM SIGIR Conference*, pages 3–12, 2014.
- [5] L. Azzopardi and M. de Rijke. Automatic construction of known-item finding test beds. In *Proceedings of SIGIR '06*, pages 603–604, 2006.
- [6] L. Azzopardi, M. de Rijke, and K. Balog. Building simulated queries for known-item topics: an analysis using six european languages. In *Proc. of the 30th annual international ACM SIGIR conference*, pages 455–462. ACM, 2007.
- [7] L. Azzopardi, D. Kelly, and K. Brennan. How query cost affects search behavior. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in IR*, pages 23–32, 2013.
- [8] L. Azzopardi and V. Vinay. Retrievability: An evaluation measure for higher order information access tasks. In *Proc. of the 17th ACM CIKM*, pages 561–570, 2008.
- [9] L. Azzopardi and G. Zuccon. An analysis of theories of search and search behavior. In *Proceedings of the 2015 International Conference on The Theory of Information Retrieval, ICTIR '15*, pages 81–90, 2015.

- [10] L. Azzopardi and G. Zuccon. An analysis of the cost and benefit of search interactions. In *Proceedings of the 2016 ACM on International Conference on the Theory of Information Retrieval, ICTIR '16*, pages 59–68, New York, NY, USA, 2016. ACM.
- [11] L. Azzopardi and G. Zuccon. *Two Scrolls or One Click: A Cost Model for Browsing Search Results*, pages 696–702. 2016.
- [12] K. Balog, L. Azzopardi, and M. de Rijke. Formal models for expert finding in enterprise corpora. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '06*, pages 43–50, 2006.
- [13] F. Baskaya, H. Keskustalo, and K. Järvelin. Time drives interaction: simulating sessions in diverse searching environments. In *Proceedings of the 35th ACM conference on research and development in information retrieval (SIGIR)*, pages 105–114, 2012.
- [14] F. Baskaya, H. Keskustalo, and K. Järvelin. Modeling behavioral factors in interactive information retrieval. In *Proceedings of the 22nd ACM international conference on Conference on information and knowledge management*, pages 2297–2302, 2013.
- [15] M. J. Bates. The design of browsing and berrypicking techniques for the online search interface. *Online Information Review*, 13(5):407–424, 1989.
- [16] U. Birchler and M. Butler. *Information economics*. Routledge, 2007.
- [17] M. D. Cooper. A cost model for evaluating information retrieval systems. *Journal of the American Society for Information Science*, pages 306–312, 1972.
- [18] S. Erdelez. Information encountering: a conceptual framework for accidental information discovery. In *Proceedings of an international conference on Information seeking in context*, pages 412–421. Taylor Graham Publishing, 1997.
- [19] R. T. Fernández, D. E. Losada, and L. A. Azzopardi. Extending the language modeling framework for sentence retrieval to include local context. *Information Retrieval*, 14(4):355–389, 2011.
- [20] F. S. Hillier and G. J. Lieberman. *Introduction to operations research*. NY, US, 2001.
- [21] P. Ingwersen and K. Järvelin. *The Turn: Integration of Information Seeking and Retrieval in Context*. Springer-Verlag New York, Inc., 2005.
- [22] K. Järvelin. Interactive relevance feedback with graded relevance and sentence extraction: simulated user experiments. In *Proceedings of the ACM conference on Information and knowledge management (CIKM)*, pages 2053–2056, 2009.
- [23] A. Kashyap, V. Hristidis, and M. Petropoulos. Facetor: cost-driven exploration of faceted query results. In *Proceedings of the 19th ACM conference on Information and knowledge management (CIKM)*, pages 719–728. ACM, 2010.
- [24] D. Kelly and L. Azzopardi. How many results per page?: A study of serp size, search behavior and user experience. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '15*, pages 183–192, 2015.
- [25] C. C. Kuhlthau. Developing a model of the library search process: Cognitive and affective aspects. *RQ*, pages 232–242, 1988.
- [26] K. G. Murty. Optimization models for decision making: Volume. *University of Michigan, Ann Arbor*, 2003.
- [27] P. Pirolli and S. Card. Information foraging. *Psychological Review*, 106:643–675, 1999.
- [28] D. M. Russell, M. J. Stefik, P. Pirolli, and S. K. Card. The cost structure of sensemaking. In *Proceedings of the INTERACT/SIGCHI*, pages 269–276, 1993.
- [29] H. A. Simon. A behavioral model of rational choice. *The quarterly journal of economics*, 69(1):99–118, 1955.
- [30] H. A. Simon. Theories of bounded rationality. *Decision and organization*, 1:161–176, 1972.
- [31] M. D. Smucker and C. L. Clarke. Time-based calibration of effectiveness measures. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in IR*, pages 95–104, 2012.
- [32] M. D. Smucker and C. L. Clarke. Modeling optimal switching behavior. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval, CHIIR '16*, pages 317–320, 2016.
- [33] M. D. Smucker and C. P. Jethani. Human performance and retrieval precision revisited. In *Proceedings of the 33rd ACM conference on research and development in information retrieval (SIGIR)*, pages 595–602, 2010.
- [34] G. J. Stigler. The economics of information. *The journal of political economy*, 69(3):213–225, 1961.
- [35] C. J. van Rijsbergen. *Information Retrieval*. Butterworth, 1979.
- [36] H. R. Varian. How to build an economic model in your spare time. *American Economist*, 41:3–10, 1997.
- [37] T. D. Wilson. Human information behavior. *Informing science*, 3(2):49–56, 2000.