University of Glasgow | School of Computing Science

The Economics of Searching

Sean Galbraith

29 March 2013

Abstract

Searching as an activity can be defined as a selection of preferences and strategies employed by a user in order to receive a particular result. By modelling this behaviour, we may construct a process through which predictions can be made regarding the decisions a user will make when faced with specific situations throughout the searching process. In particular, we study the user's response to relative costs associated with the activities of reading results and engaging in new queries given some information regarding the overall precision of the query set. Information Foraging Theory proposes that users will attempt to maximise the value of information they will receive given certain characteristics of search environment, including the cost associated with certain actions and, as we will focus on, any knowledge the user has about the likelihood of receiving highly relevant information.

This report aims to quantify these values over three separate scenarios - low, medium and high yield queries - by use of a high level abstraction of the searching process in the form of a simple game. Through this model of behaviour, predictions may be made as to the stopping distance of users on a particular query given certain variables, namely its yield and the user's level of experience.

Contents

Chapter 1

Introduction

1.1 Motivation and Context

1.1.1 An Optimal Approach to Searching

For any given set of queries and associated documents, there is an optimal path through which a user must tread in order to achieve the highest possible gain - that is, return the greatest number of highly relevant results. With no restrictions, this path is a simple line straight through all queries and documents - there is no need for the user to stop at any point. However, should restrictions come into play, the user must make a decision as to how they proceed in the searching process. The optimal path, in this case, is defined as the correct sequence of result / query decisions made which expend all resources to return the highest gain.

1.1.2 Reaching the Optimal Approach

Given that there exists an optimal approach, the next topic to be studied relates to the user and their tendency to that particular strategy. According to Information Foraging Theory suggested by [**?**], users will naturally adapt their strategy in order to achieve a result closer to this optimal approach. By studying a users' reaction over a period of time, we can test this hypothesis while extending it to look at user behaviour under specific scenarios to note the effect of a specific variable, precision, on this adaptation of behaviour.

1.1.3 Building a Better Search

Search engines and other providers currently present users with information in a manner which lacks optimal efficiency, as shown by the lack of users ability to consistently procure all relevant documents while expending the minimal amount of effort within this experiment and others such as [[LEIF TREC stuff + others]]. Therefore, if the accuracy and reliability of the prediction is of a sufficient quality, the efficiency of this transaction may be vastly improve by culling the data sent to a minimal amount over the proposed stopping limit.

1.2 Research Questions

The objective of this paper is to leverage the ideas suggested by Information Foraging Theory to map user behaviour over differing levels of precision amongst queries. With this, we will test if the data can be used to provide accurate stopping distances in an experimental context. In order to better understand the specifics, it can be broken down into three sub-questions as follows:

1.2.1 User Strategy

The first and most vital point to be addressed is the question: 'Do users really employ a strategy?' In order for the basic tenants of Information Foraging Theory to hold, users must interpret information they are given (whether in the form of knowledge or constraints) as to alter their behaviour towards the path of what they think is the most optimal solution. By varying this information we may in effect pose this question directly to the users, in order to see if they truly react to it. Should the answer to this question prove to be positive, we may move on to a more accurate analysis of the user strategy.

1.2.2 Predictability

Continuing from the previous point, we may pose a follow-up question: 'Can this strategy be predicted?' If it is true that users are employing strategies to attempt to 'game' their way to the most optimal outcome, then it may be possible to predict whether a user will perform one of several actions at any given point. With a high level approach involving two possible decisions related to the searching process, reading a further result or switching to a new query, we may measure strategy by the 'stopping distance' at which a user ceases to read results of one query and moves on to the next. With the result of this question, the research will provide the information sought as the original motivation: An accurate level of information which must be provided and superfluous information which, should the user employ this strategy, not be required.

1.2.3 Learning Effect

A final question we will address in passing is 'How quickly, if ever, do users move towards more optimal strategies?' Over repeated attempts at reaching a particular strategy, a user has gained more and more knowledge of the optimal path on each occasion, allowing them a better chance of reaching it. This introduces a bias which may interfere with the previous questions, and as such the design of the experiment takes this into account.

1.3 Summary of the Contributions

Throughout the process of developing this report, a number of developments were incurred. The final outcome of course is the conclusion of the research, supported by the production of various tools in order to aid this process, outlined as follows:

1.3.1 An Abstract Searching Game

The major tool used throughout the experimental stage of this process is presented as a high level abstraction of the searching process in the form of a game. Written in Python with a purely textual interface, the game provides a user with the pure, abstract searching process and a simple process whereby two decisions can be chosen between, eliminating any outside bias which may be prevalent in the process of a full search engine. The game is reliant on its input dataset and produces an output of results including actual stopping distances, which can be used to measure their relation to the proposed stopping distance of a particular input dataset.

1.3.2 A 'Yield' Based Dataset

A dataset of queries and results of said queries constructed in order to provide three specific scenarios. These three scenarios relate to the precision of the queries contained within, and are defined as such: High Yield Queries: Queries with a relatively high precision. Medium Yield Queries: Queries with a middle-level precision. Low Yield Queries: Queries with a low precision.

1.3.3 A System of Prediction

The hypothesis of our research: That given a certain dataset, stopping distances may be predicted. By using the high, medium and low yield queries as our base, we can begin a process whereby we construct and estimate stopping distances based upon the nature of those yields in relation to information foraging theory.

1.3.4 Experimental Confirmation

With our experiment process in the form of the game, our experimental dataset of yield-based queries and our the outcomes of our process of estimating predictions, experiments can be run allowing subjects to make their way through the game while it records their actions as the result of a query which they stop at. This provides a result which may be directly compared with the estimated prediction.

1.3.5 Resultant Conclusion

A report and analysis will therefore be constructed in conclusion in order to show the extent to which the experiment confirms our hypothesis.

1.4 The Structure or Outline of the Report

Chapter 2

Related Work or Background

Summarising a paper: (1) Context, (2) Problem, (3) Solution, (4) Evaluation, (5) Impact.

Places to go: ACM Portal, Google Scholar, Citeseer

Journals: IPM, JIR, ACM TOIS Conferences: ACM SIGIR, ACM CHI, ACM CIKM, European Conference in IR (ECIR), Information Interaction in Context (IIiX), etc.

Types or Styles of Papers: Theoretical, Empirical, Conceptual, Applications Based.

It is your job to add value and show how the background work relates to your project

- What is it about?
- Why would i read it? What is of value in it?
- What are the main contributions in the paper?
- What are the main issues in the paper?
- What are the advantages and disadvantages of the approach/solution proposed?
- What are the limitations of the work?
- What does the paper claims does the paper make? And are they supported?
- What do other people think of the paper? Who has caned it?
- Consider whether the paper is seminal or delta?
- What did you learn from this paper?
- Who else has done work in this area?
- How does this work stand out?
- How does it relate to the research questions?

Chapter 3

Methodology

3.1 An Experimental Approach

In order to test if the theory of our research holds up in a real-world environment, an experiment had to be run as a test case and source of evaluation. The experiment will test our hypothesis, a set of constructed queries and estimated stopping distances against the experimental output - a user's real stopping distance when faced with those same queries.

3.2 Outlining the Experiment

3.2.1 Aims

1. Provide an abstracted environment for a user to perform search-like functions.
2. Construct an estimation of user behaviour using prior research on Information Foraging Theory, query precision and stopping distances.
3. Observe the behaviour of a user when faced with theories of varying yield.
4. Record certain aspects of this behaviour, with a focus on stopping distance.
5. Compare the expected behaviour with the actual behaviour.
6. Report the outcome of our hypothesis test.

Throughout the remainder of this section, we will explain in depth the process through which these methods were chosen along with the design decisions taken in order to include them in the experiment in the most efficient manner possible.

3.3 Abstraction

3.4 Designing a Game

3.5 User Interface

3.6 Experimental Procedure

3.7 Query Selection

3.8 User Explanation

3.9 Learning Through Repetition

3.10 Storing Information

3.11 Collating Results

Chapter 4

Results and Analysis

Chapter 5

Discussion and Conclusion

5.1 Acknowledgements

I would like to thank ...