SciNet: A System for Browsing Scientific Literature through Keyword Manipulation

Dorota Głowacka
Department of Computer Science
University of Helsinki
glowacka@cs.helsinki.fi

Tuukka Ruotsalo
Helsinki Institute for Information Technology, HIIT
Aalto University
tuukka.ruotsalo@aalto.fi

Ksenia Konyushkova
Department of Computer Science
University of Helsinki
ksenia.konyushkova@helsinki.fi

Kumaripaba Athukorala
Department of Computer Science
University of Helsinki
kumaripaba.athukorala@hiit.fi

Samuel Kaski
Helsinki Institute for Information Technology, HIIT
Aalto University
samuel.kaski@aalto.fi

Giulio Jacucci
Department of Computer Science
University of Helsinki
giulio.jacucci@helsinki.fi

ABSTRACT

Techniques for both exploratory and known item search tend to direct only to more specific subtopics or individual documents, as opposed to allowing directing the exploration of the information space. We present SciNet, an interactive information retrieval system that combines Reinforcement Learning techniques along with a novel user interface design to allow active engagement of users in directing the search. Users can directly manipulate document features (keywords) to indicate their interests and Reinforcement Learning is used to model the user by allowing the system to trade off between exploration and exploitation. This gives users the opportunity to more effectively direct their search.

Author Keywords

Adaptive Interfaces, Information Retrieval, Datamining and Machine Learning, Recommender/Filtering Systems

ACM Classification Keywords

H.5.2. User Interfaces: User-centered design

INTRODUCTION

In a typical interaction with an information retrieval system, the user enters a specific query, investigates the results returned by the search engine, and alters the query based on the results to direct the search to a more specific subtopic or drift to an alternative direction. Thus, users frequently have to carefully investigate the results to be able to reformulate their query. Recent studies have shown that by modeling the search context, a system can provide much richer information about the search intention, limit the number of alternatives users need to select from to direct their search, and automate the tedious query reformulation process [5]. Reinforcement learning (RL) is a promising approach that allows the system to trade between exploitation (moving towards more specific subtopics) and exploration (going towards alternative topics), and has been shown to be helpful in information retrieval [2, 4], recommender systems [2] or ads placing [3]. However, most systems that employ RL rely on collecting information on users' interests over a pro–longed period of time, while in a typical search scenario users are more interested in the overall improvement of the search results within a given search session rather than hypothetical future search sessions.

We propose that better support for exploration can be provided through learning from feedback on higher level representations of the data, such as keywords extracted from documents. This allows users to direct their search using the offered keyword without getting trapped into a context, or having to provide tedious document–level relevance feedback, nor relying on implicit feedback mechanisms that may take long to converge. The learning mechanism also predicts the visualized keywords based on the search session context and allows users to direct the search rather than re-ranking as in traditional personalization. The resulting system couples advanced machine learning techniques with information visualization and interaction to boost exploratory search and allows the users to actively engage in the exploration.

SYSTEM OVERVIEW

The primary goal of SciNet is to assist scientists in finding and exploring relevant literature on a given research topic quickly and effectively. The visual search interface is presented in Figure 1. The user assigns relevance scores to the displayed keywords by moving them within the exploratory view (the circle on the left-hand side) and thus directs the search according to her interest. The inbuilt RL mechanism helps SciNet to form a model of user's interests and suggest appropriate keywords in future search iteration.

The search starts with the user typing in a query, which results

IUI'13 Companion, March 19–22, 2013, Santa Monica, CA, USA.
ACM 978-1-4503-1966-9/13/03.

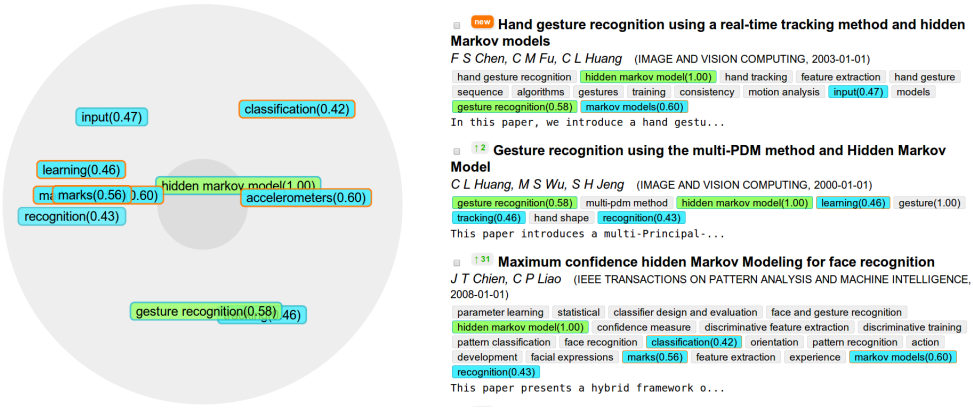

Figure 1. An example of a search session using SciNet. The keywords displayed in the exploratory view (left) offer the user the option to continue the exploration in new directions, such as "learning" or "classification". Keywords explicitly manipulated by the user in the previous iteration are colored in green. Keywords that never appeared in previous iterations can be distinguished by their orange borders. The document list has both new documents (labeled "new") and documents whose rank increased from the previous iteration.

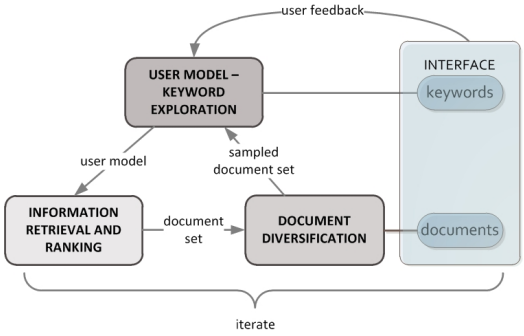

Figure 2. Overview of data flow in the SciNet system

in a set of keywords being displayed in the exploratory view and a ranked list of articles being displayed on the right hand-side of the screen. The user can manipulate the keywords in the exploratory view to indicate how relevant they are to her search: the closer to the center a given keyword is, the more relevant it is to the user. The user can manipulate as many keywords as she likes as well as drag keywords from underneath the displayed articles into the exploratory view. After each iteration, the user model is updated and new keywords and articles are displayed. The search continues until the user is satisfied with the results.

The data flow from the system's perspective is illustrated in Figure 2. The three main blocks of the system are responsible for (1) retrieval and ranking of documents, (2) user modeling and keyword exploration, and (3) document diversification.

The initial set of documents and their rankings are obtained through the Information Retrieval and Ranking component. Having received feedback on keywords, the system enters the exploratory loop. The explicit user feedback is sent to the Keywords Exploration and the Document Diversification components. The Keywords Exploration module implements user model estimation using RL techniques. The user model is a representation of the system's belief about the user's informational need at the current iteration of retrieval. The component receives feedback from the user and produces a list of keywords with weights which are passed on to the Information Retrieval and Ranking module, which predicts a new set of documents for the new search iteration. Thus, the dataset in the system is not static and it changes at every iteration based on the present, best estimation of the user model.

The Document Diversification module is responsible for determining the set and order of documents that are passed on to the Interface. The module uses exploration–exploitation techniques to sample a set of documents to display to the user, while keeping the ranking obtained from Information Retrieval and Ranking component. The new set of documents is used in Keywords Exploration component to capture dependencies between keywords.

The user model is visualized in the exploratory view, which allows the user to give feedback to the system through keyword manipulation. A list of articles is also presented to the user. The system gets new feedback from the user and continues in the iterative feedback loop. A detailed description of each component and SciNet evaluation are presented in [1].

ACKNOWLEDGMENTS

The data used in the system is derived from the Web of Science prepared by THOMSON REUTERS, Inc., Philadelphia, Pennsylvania, USA; the Digital Library of the Association of Computing Machinery (ACM); the Digital Library of Institute of Electrical and Electronics Engineers (IEEE), and the Digital Library of Springer.

REFERENCES

1. Głowacka, D., Ruotsalo, T., Konyushkova, K., Athukorala, K., Kaski, S., and Jacucci, G. Directing exploratory search: Reinforcement learning from user interactions with keywords. In *Proc. of IUI* (2013).
2. Joachims, T., Freitag, D., and Mitchell, T. Webwatcher: A tour guide for the world wide web. In *Proc. of IJCAI* (1997).
3. Li, L., Chu, W., Langford, J., and Schapire, R. A contextual-bandit approach to personalized news article recommendation. In *Proc. of WWW* (2010).
4. Seo, Y., and Zhang, B. A reinforcement learning agent for personalized information filtering. In *Proc. of IUI* (2000), 248–251.
5. White, R. W., Bennett, P. N., and Dumais, S. T. Predicting short-term interests using activity-based search context. In *Proc. of CIKM* (2010).