

# User models from implicit feedback for proactive information retrieval

Samuel Kaski<sup>1</sup>, Petri Myllymäki<sup>2</sup>, and Ilpo Kojo<sup>3</sup>

<sup>1</sup> University of Helsinki, Department of Computer Science, and Helsinki University of Technology, Neural Networks Research Centre

<sup>2</sup> Helsinki Institute for Information Technology, University of Helsinki and Helsinki University of Technology

<sup>3</sup> Helsinki School of Economics, Center for Knowledge and Innovation Research  
<http://www.cis.hut.fi/projects/mi/prima>

**Abstract.** Our research consortium develops user modeling methods for proactive applications. In this project we use machine learning methods for predicting users' preferences from implicit relevance feedback. Our prototype application is information retrieval, where the feedback signal is measured from eye movements or user's behavior. Relevance of a read text is extracted from the feedback signal with models learned from a collected data set. Since it is hard to define relevance in general, we have constructed an experimental setting where relevance is known a priori.

## 1 Introduction

Successful proactivity, i.e. anticipation, in varying contexts requires generalization from past experience. Generalization, on its part, requires suitable powerful (stochastic) models and collection of data about relevant past history to learn the models.

Information retrieval is an example of a task which would benefit from proactivity: the user has a goal—to find relevant pieces of information—which is not directly observable to the system and needs to be inferred from the actions. Information retrieval applications appear in various contexts from traditional searching of text or multimedia documents to computerized manuals and help systems. Analogously, a user entering a room, for instance a shop, in which multiple actions are possible, could be assisted in choosing from a set of “relevant” actions. Some experimental information retrieval systems take the user's goals into account, but they rely heavily on explicit input. To the extent that implicit information is used, it has been restricted to heuristic estimation of relevance from navigation data.

The overall interest profiles of user populations have been modeled in collaborative filtering systems recommending potentially interesting items for groups of similar users. Information about the current interest of a user, on the other hand, has been mostly acquired as explicit relevance feedback, by displaying the user a set of retrieved items and asking which are relevant. The problem with both approaches is that interactive collection of relevance information is laborious, and

it becomes outdated quickly. Moreover, only a small amount of information per item can be collected. Hence, methods that would infer multi-faceted relevance from the natural actions of users would be extremely valuable.

Inferring the goals of users requires measuring signals that (i) contain a sufficient amount of information about the intentions of the user, (ii) have a sufficiently high “relevant-signal-to-noise ratio” that at least some aspects of the intentions can be extracted and modeled, and (iii) can be measured feasibly also in practice. We use *eye movement* signals that exhibit both voluntary and involuntary signs of the interests and intentions.

## 2 Status

We have developed the first versions of machine learning methods for inferring relevance from eye movements [1] and studied coupling of text content to the task of estimating relevance [2]. We are currently working on an information retrieval application, on incorporating other forms of implicit relevance feedback, and on studying implications of the work in perceptual psychology.

## 3 Eye movement challenge

We have started a competition, with deadline in September 2005, on predicting relevance of sentences from eye movement data. The data has been collected in a controlled experimental setting in which the relevance is known, which makes machine learning methods directly applicable.

The challenge is at <http://www.cis.hut.fi/eyechallenge2005/>; participation is open to all.

## Acknowledgements

The project is funded by the Academy of Finland, decision number 122209, and by the IST Programme of the European Community, under the PASCAL Network of Excellence, IST-2002-506778. This publication only reflects the authors’ views. The authors would like to thank all other people in the PRIMA project and acknowledge that access rights to the data sets and other materials produced in the PRIMA project are restricted due to other commitments.

## References

1. Salojärvi, J., Puolamäki, K., Kaski, S.: Relevance feedback from eye movements for proactive information retrieval. In Heikkilä, J., Pietikäinen, M., Silvén, O., eds.: workshop on Processing Sensory Information for Proactive Systems (PSIPS 2004), Oulu, Finland (2004)
2. Savia, E., Kaski, S., Tuulos, V., Myllymäki, P.: On text-based estimation of document relevance. In: Proc. IJCNN’04. (2004) 3275–3280