Chapter 9

Proactive information retrieval

Samuel Kaski, Kai Puolamäki, Antti Ajanki, Jarkko Salojärvi

9.1   Introduction

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

The goal of the PROACT project is to build statistical machine learning models that learn from the actions of people to model their intentions and actions. The models are used for disambiguating the users' vague commands and anticipating their actions.

Our application area is information retrieval, where we investigate to what extent the laborious explicit relevance feedback can be complemented or even replaced by implicit feedback derived from patterns of eye fixations and movements that exhibit both voluntary and involuntary signs of users intentions. Inference is supported by models of document collections and interest patterns of users.

The PROACT project has been done in close collaboration with researchers in the European Union's Pascal Network of Excellence within a Pump Priming Programme (2005–2007); the collaborators are from University of Helsinki, University of Southampton and University College London. The project continues in a STREP project PinView from 2008 onwards.

9.2   Implicit queries from eye movements

Eye movements measured during reading are a promising new source of implicit feedback. During complex tasks such as reading, attention approximately lies on the location of the reader's gaze. Therefore the eye movements should contain information on the reader's interests. Inferring interest of the user from a reading pattern is difficult however, since the signal is complex and very noisy, and since interestingness or relevance is higly subjective and thus hard to define. We have earlier developed machine learning and signal processing methods for this task, and hosted a research challenge where the task was to predict relevance from eye movement patterns [1].

The motivation for the next stage of the research was that formulating a good query in a web search engine, for example, is known to be difficult. Implicit feedback collected by observing user's behavior might reveal the true interest of the user without the need to explicitely label the documents as relevant or not relevant. We performed a feasibility study in which we used eye movements to formulate a query that reflects user's interest while he was reading [2].

We constructed a controlled experimental setting in which it is known which documents are relevant. The users read short documents searching for the ones related to a topic that was given to them beforehand. The eye movements were recorded during reading. We trained a regressor that predicts how relevant a term is for user's current query given the eye movement measurements on that term. This regressor can then be applied to new topics, with no training data available, to estimate relevance of words.

The learned model was then used to infer relevant query terms based on eye movement recorded while the user was performing a new search task, where the true query was unknown. The inferred query terms can be used to retrieve and suggest new documents that might be important to user's information need.

A SVM model that uses eye movements and textual features outperformed a similar model without the eye movement features. This indicates that eye movements contain exploitable information about relevance in information retrieval tasks.
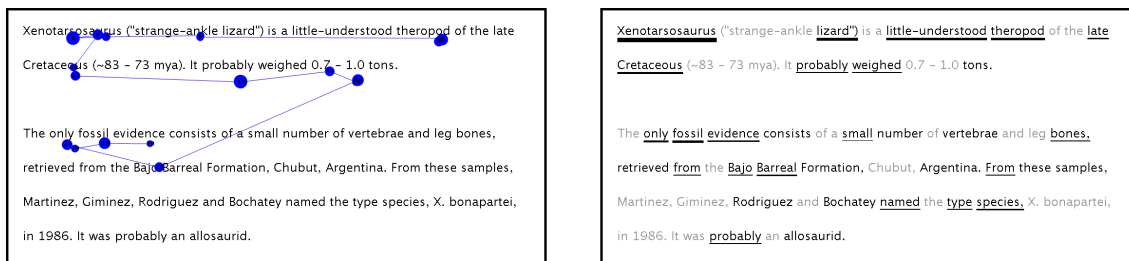

Figure 9.1: Left: A sample eye movement pattern of a test subject during reading a document. Right: The term weights depicted on the same document inferred from the eye movements of all test subjects who were searching for information about dinosaurs. The magnitudes of the weights are depicted as the thickness of the underlining.

References

[1] Kai Puolamäki and Samuel Kaski, editors. *Proceedings of the NIPS 2005 Workshop on Machine Learning for Implicit Feedback and User Modeling.* Helsinki University of Technology, Espoo, Finland, 2006.

[2] David R. Hardoon, John Shawe-Taylor, Antti Ajanki, Kai Puolamäki, and Samuel Kaski. Information Retrieval by Inferring Implicit Queries from Eye Movements. In *Proceedings of the 11th International Conference on International Conference on Artificial Intelligence and Statistics*. San Juan, Puerto Rico, 2007.