Relevance feedback from eye movements for proactive information retrieval

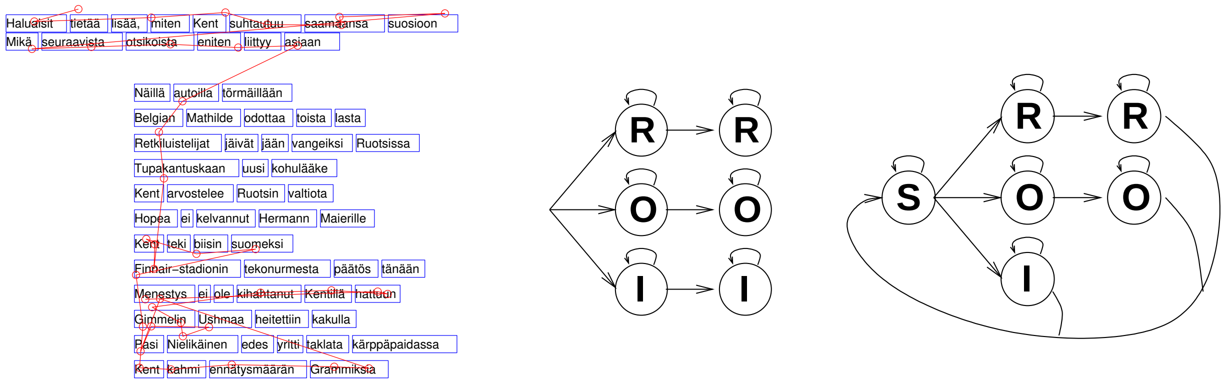
Jarkko Salojärvi, Kai Puolamäki, and Samuel Kaski, Neural Networks Research Centre, Helsinki University of Technology, P.O. Box 9800, FIN-02015 HUT, Finland. Email: {forename.surname}@hut.fi

Topics: 1. User modeling and personalization, 2. Multimodal and perceptual user interfaces

Introduction. This on-going work is part of a larger project that aims at complementing or even replacing the laborious explicit relevance feedback in information retrieval by implicit feedback. Technology for measuring eye movements begins to be mature enough, and the movements contain rich information about the attention and interest patterns of the user. The problem is that the signal is very noisy and the correspondence of the eye fixation patterns to user's attention is not always one-to-one.

Eye movements have earlier been used as alternative input devices in human-computer interfaces, and recently in a proactive dictionary which becomes automatically activated [2]. To our knowledge, they have not been used in information retrieval before. We have implemented a pilot system and use it for developing models for detecting cues of relevance from eye movements. We started by preliminary experiments on whether simple HMM-based models are capable of discriminating relevant texts.

Setting. The task is to predict, from the pattern of eye movements, whether a document title is relevant to the user. In order to apply machine learning methods we collected a set of patterns in a controlled experimental setting where relevance was known (an example in Fig. 1**a**). The user was asked a question and some of the titles on the display were known to be relevant (class "R"), and one of them contained the answer (class "O"). The rest were irrelevant (class "I").



a **b** **c**

Figure 1: **a** Sample eye movement pattern. Lines connect successive fixations, denoted by circles. Each line contains one title (in Finnish). **b** Two-stage HMMs of reading patterns. **c** Combined HMM. Each state emits a three-dimensional feature vector for a word, and transitions occur between words. Descriptive titles for the states: S: Scanning mode, R: relevant title, O: title contains the answer to the question, I: irrelevant title.

Feature extraction. In psychological studies of reading it is common to segment the eye movement signal into a sequence of fixations and saccades, and compute summary measures of them for each word. We started with a common set of 21 such features [3], and reduced the dimensionality by Automatic Relevance Determination (`http://www.cs.toronto.edu/~radford/fbm.software.html`). The resulting features measured (1) fixation duration, (2) how many fixations per word, and (3) where the attention moves after the current word.

Table 1: Accuracy of classifying test titles to the three classes. The Combined HMM is significantly better than the next-best alternative (McNemar's test).

Method	Dumb	LDA	2-state HMMs	Combined HMM
Accuracy (%)	63.2	69.2	71.3	**75.8**

Modeling of reading patterns. The sequential nature of the reading process is taken into account by a Hidden Markov Model (HMM). HMMs have been used for segmenting the low-level eye movement signal to detect focus of attention (see [6]) and for implementing (fixed) models of cognitive processing [4]. We learn HMMs from word-level eye movement data to discriminate between the three classes.

We started by comparing the performance of three simple models to assess whether taking the sequential nature of the signal into account improves prediction. The first model is a simple linear predictor (linear discriminant analysis, LDA) that uses the average feature vector computed over each title. The second model is the simplest possible way of taking sequence characteristics into account: a simple two-state HMM is optimized for each class (Fig. 1**b**). The output is the MAP prediction.

The user typically alternates between two (or more) different subtasks, searching and reading, while viewing the screen. The third model is the simplest possible way of taking these alternations into account. The model now contains the searching state ("S" in Fig. 1**c**) as well, and it is optimized in a simple and fast way: The branches are fixed to values estimated in stage 2 above, and only the rest of the model is estimated from the complete data ("Combined HMM"). Model complexity was controlled by reducing the number of states in the "I"-branch.

Taking the sequence into account improves discrimination accuracy slightly (Table 1), and adding the strategy state further improves the model.

Discussion. The first experiments are promising in that even the simple HMM structures with partly heuristic training procedures clearly improve on the results of the non-sequence model. The remaining main questions are (1) what kind of a model structure to use, (2) how to best optimize them in a discriminative way, and (3) how to use the models to discovering cues about relevance.

For optimization we will next investigate the discriminative learning approach that maximizes the mutual information between the (estimated) labels emitted by the model and the correct labels [5]. The HMM parameters are optimized by an extended Baum-Welch (EM) algorithm.

The present models work on word-level eye-movement data. They will be later complemented with lower-level generative models of the eye movement signals (cf. [1]) and combined with models of document textual content.

References

[1] R. Engbert, A. Longtin, and R. Kliegl. A dynamical model of saccde generation in reading based on spatially distributed lexical processing. *Vision Research*, 42:621–636, 2002.

[2] A. Hyrskykari, P. Majaranta, and K.-J. Räihä. Proactive response to eye movements. In *Proc. INTERACT'03*. To appear.

[3] K. Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124:372–422, 1998.

[4] D. D. Salvucci and J. R. Anderson. Automated eye-movement protocol analysis. *Human-Computer Interaction*, 16:39–86, 2001.

[5] R. Schlüter and W. Macherey. Comparison of discriminative training criteria. In *Proc. ICASSP'98*, pages 493–496. 1998.

[6] C. Yu and D. H. Ballard. A multimodal learning interface for grounding spoken language in sensory perceptions. In *Proc. ICMI'03*. To appear.