RELEVANCE FEEDBACK FROM EYE MOVEMENTS FOR PROACTIVE INFORMATION RETRIEVAL

Jarkko Salojärvi [1], *Kai Puolamäki* [1], *Samuel Kaski* [1,2]

[1]Helsinki University of Technology
Neural Networks Research Centre
P.O. Box 5400, FIN-02015 HUT, Finland

[2]Department of Computer Science
P.O. Box 26, FIN-00014 University of Helsinki, Finland

ABSTRACT

We study whether it is possible to infer from eye movements measured during reading what is relevant for the user in an information retrieval task. Inference is made using hidden Markov and discriminative hidden Markov models. The result of this feasibility study is that prediction of relevance is possible to a certain extent, and models benefit from taking into account the time series nature of the data.

1. MOTIVATION

Proactive computing applications try to predict the needs of the user and adapt their own behavior accordingly [1]. As a concrete example, in information retrieval (IR) not all people find the same articles suggested by a search engine to be relevant, at least not to the same degree. In order to tune an IR application to find documents more closely matching the preferences of an individual user, a source of feedback is needed. The usual way would be to ask after every document whether the user found it relevant or not, and learn the user's preferences from the answers. However, this kind of approach to get explicit feedback is often considered to be laborious.

The relevance of a document can alternatively be inferred from implicit feedback. The idea is to obtain information on preferences unobtrusively, by monitoring users' natural interactions with the system [2]. Traditionally implicit feedback has been derived from document reading time, or by monitoring saving, printing, or selecting of documents. We suggest monitoring eye movements during information retrieval as a source of implicit feedback. The technology for measuring eye movements begins to be mature enough, and it is a fact that the movements contain rich information about the attention and interest patterns of the user [3]. The problem is that the signal is very noisy and the correspondence of the eye movement patterns to user's attention is sometimes ambiguous.

In order to determine whether relevance can be inferred from eye movements in an information retrieval task, we devised a controlled experimental setup where the relevant items are known, and then measured the eye movements of test subjects while they were carrying out the experiment. Initial data exploration on eye movement data [4] verified that the eye movement data is not too noisy for inferring relevance. In this paper we explore whether it is possible to find more fine-grained cues of relevance using more sophisticated models. We report results of a feasibility study where hidden Markov models (HMMs) are used to discriminate relevant texts using standard eye movement features reported in psychological literature.

The research questions which this feasibility study was aimed at answering are: (1) do the models benefit from the time series nature of the data, (2) do discriminative models improve performance, (3) does modeling of the global scanning behavior help in predicting relevance, and (4) is it possible to discover reading strategies of the user with different HMM structures.

2. EYE MOVEMENTS AS A SOURCE OF FEEDBACK

2.1. Physiology

The eye movement pattern consists of rapid eye movements, *saccades*, followed by *fixations* during which the eye is fairly motionless. The reason for the pattern lies in the anatomy of the retina; due to rapidly decreasing visual cell density towards periphery, accurate viewing is possible only in the central *fovea* area where the density is high. The area spans only 1–2 degrees of visual angle. Therefore, detailed inspection of a scene has to be performed in a sequence of saccades and fixations, often referred to as a *scanpath*. An example of a scanpath during an information retrieval task is shown in Figure 1. The duration of a fixation is correlated with the complexity of the object currently under inspection. During reading this complexity is associated with the frequency of occurrence of the words in general, and with how predictable the word is from its context [3]. Naturally there are other factors affecting the reading pattern as well, such as different reading strategies and the mental state of the reader.

2.2. Earlier work

In psychology, study of eye movements as an indicator of low-level cognitive processes is a well-established research area [3]. However, fewer attempts to infer higher order cognitive processes from eye movements can be found. This is mainly due to the fact that it is extremely difficult to construct a controlled experiment where only one cognitive aspect affecting the eye movements can be measured. An example of such an experiment is [5], where correlation between the pupil size and difficulty of processing a sentence was reported.

Use of eye movements as a source of implicit feedback is a relatively new concept. Eye movements have earlier been utilized as alternative input devices for either pointing at icons or typing text in human-computer interfaces (the most recent application being [6]). The first application where user interest was inferred from eye movements was an interactive story teller [7]. The story told by the application concentrated more on items that the user was gazing at on a display. Rudimentary relevance determination is needed also in [8], where a proactive translator is activated if the reader encounters a word which she has difficulties (these are inferred from eye movements) in understanding. A prototype attentive agent application (Simple User Interest Tracker, Suitor) is introduced in [9, 10]. The application monitors eye movements during browsing of web pages in order to determine whether the user is reading or just browsing. If reading is detected, the document is defined relevant, and more information on the topic is sought and displayed. The rules for inferring whether the user is reading are determined heuristically [11].

3. EXPERIMENTS

Since relevance of a document is subjective in general, we designed a controlled experiment with known relevance.

3.1. Experimental setup

In the experiment, the subject was instructed to find an answer to a question from a list of twelve titles (sentences). Eight of the titles were known to be irrelevant (I), three relevant for the question (R), and one contained the correct answer (C). Each of the three test subjects carried out 15 assignments. Eye movements were measured with a head-mounted eye tracker consisting of a helmet with two cameras; one monitored the eye and the other one the visual field of the subject (see Figure 1). The raw eye movement data (x and y coordinates of where the subject was looking, measured with a sampling rate of 50 Hz) was then segmented into a sequence of fixations and saccades by software from equipment manufacturer.

(a)

(b)

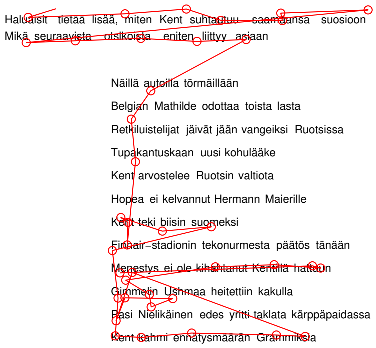

Fig. 1. (a): Eye movements were measured with a head-mounted eye tracker (iView from SensoMotoric Instruments GmbH, Germany). (b): Sample eye movement pattern during an information retrieval task. Lines connect successive fixations, denoted by circles. Each line contains one title (in Finnish).

3.2. Feature extraction

In psychological studies of reading, summary measures of the segmented eye movement signal are computed for each word. In preprocessing, each fixation is first heuristically assigned to the nearest word. Using this assignment, a common set of 21 such features found from literature [3, 12] were computed. Then the dimensionality was reduced by a Bayesian multilayer perceptron (MLP) having an Automatic Relevance Determination (ARD) prior[1]. The features

[1] software package available from
http://www.cs.toronto.edu/~radford/fbm.software.html.

resulting in the best classification accuracy were:

1. One or many fixations (binomial).

2. Logarithm of total fixation duration (assumed Gaussian).

3. Reading behavior (multinomial): skip next word, go back to already read words, read next word, jump to an unread line, or last fixation in an assignment.

The features were computed for each word along the eye movement trajectory. The whole trajectory was segmented to sequences occurring on the same title, and a label was assigned to each sequence according to the class of the title. The goal of this work was to try to predict the known relevance of the title from the measures.

4. MODELS

In an earlier exploration on eye movement data [4], we analyzed title-spesific averages of eye movement measures computed for each word in the title. In this paper we take a closer look at the data with models that take into account the time series nature of the eye movement data. We estimate hidden Markov models from word-level eye movement data to discriminate between the three classes of titles. Ordinary HMMs and discriminative HMMs are applied, optimized by maximum likelihood (ML).

4.1. Hidden Markov Models

In order to explain user behavior, the sequential nature of the reading process has to be modelled. The most common approach to model sequential data is using hidden Markov models. In eye movement research, hidden Markov models have earlier been used for segmenting the low-level eye movement signal to detect focus of attention (see [13]) and for implementing (fixed) models of cognitive processing [14], such as pilot attention patterns [15].

Hidden Markov models optimize the log-likelihood of the data Y given the model and its parameters Θ, that is, $\log p(Y|\Theta)$. The goal is to optimize the parameters of the model so that the distribution of the data is expressed as accurately as possible. HMMs are *generative models*; they attempt to describe the process of how the data is being generated. Therefore they can be said to *emit* (produce) observations.

Long-range time dependencies within the data are taken into account by adding hidden states to the model. The changes in the distributions of the emitted observations are associated with transitions between hidden states. The transitions (as well as the observation distributions) are modelled probabilistically. There exists a well-known algorithm for learning the HMMs, namely the Baum-Welch (BW) algorithm, if all the probabilities within the model are expressed using distributions which are within the exponential family [16]. Baum-Welch algorithm is a special case of Expectation-Maximization (EM) algorithm and it can be proven to converge to a local optimum.

4.1.1. Simple Hidden Markov Model for Each Class

The simplest model that takes the sequential nature of data into account is a two-state HMM. We optimized one model individually for each class (see Figure 2). In a prediction task the likelihood of each model is multiplied by the prior information on the proportions of different classes in the data. As an output we get the maximum a posteriori prediction.

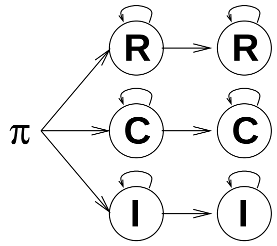

Fig. 2. A simple two-state hidden Markov model was optimized for each of the classes. $\pi = \pi_{\{R,C,I\}}$ is the prior probability of the classes.

4.1.2. Global Hidden Markov Model

During an IR task, the user typically alternates between two (or more) different subtasks, searching and reading. One possible way of taking these alternations into account is to construct a global HMM using the whole eye movement trajectory during the task, without segmenting it to title-specific sequences.

The simplest modification to our model is to combine all the individual HMMs and add a searching state ("S" for scanning in Fig. 3). When comparing the models, their complexity should be equal. This was enforced by reducing the number of states in the "I"-branch.

There are no known EM-type algorithms for optimizing this kind of a model in a way that best discriminates between classes. Therefore, in order to test whether the model is feasible, we implemented an ad hoc training method using ordinary Baum-Welch. The model is trained accordingly: each branch ("R", "C", and "I") is first optimized with BW using sequences having a class label associated with that branch. Then we optimize the whole model using complete eye movement trajectories during assignments (that is,

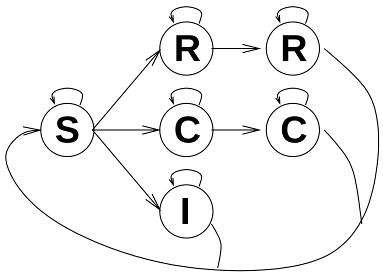

Fig. 3. The topology of a global hidden Markov model.

without segmenting them to title-specific sequences), keeping emission distributions of each of the branches "R", "C", and "I" fixed.

4.2. Discriminative Hidden Markov Models

The goal of discriminative modeling differs from conventional Maximum Likelihood. We want to optimize the discriminative power of the model, that is, to predict the class of the data sequence *given* the observations. Discriminative models concentrate more on modeling the class distributions than the whole distribution of the data. Current state-of-the-art HMMs used for speech recognition are discriminative. Discriminative training of HMMs is carried out by having certain "correct" hidden state sequence(s) in the model to always correspond to a certain class, and then maximizing the likelihood of the "correct" state sequence for the teaching data, versus all the other possible state sequences in the model [17, 18]. Such methods have not been previously applied to eye movement data.

In our setup, we want to predict the relevance $B = \{I, R, C\}$ of a document, given the observed eye movements Y. Formally, we optimize $\log p(B|Y, \Theta)$. The parameters of the discriminative HMM can be optimized with an extended Baum-Welch (EBW) algorithm, which is a modification of the original BW algorithm (derivation of the algorithm can be found for example in [19]). Optimization of the conditional log-likelihood $\log p(B|Y, \Theta)$ can be shown to be asymptotically equivalent to the conditional entropy of the relevance measures given the observations, which in turn is closely associated with mutual information.

4.2.1. Discriminative Chain of Hidden Markov Models

The main difficulty in the information retrieval setup is that relevance is associated only with titles, not with words in a title. For example, there are words in titles which are not needed in making the decision on whether the title is relevant or not. There could be many such non-relevant words in a sentence, and possibly only one word which is highly relevant. The situation thus resembles the setup in reinforcement learning: the reward (classification result) is only known in the end, and there are several ways to end in a correct classification. For the same reason, discriminative training for a global HMM as presented in Figure 3 is difficult, since there are a multitude of different paths in the model that can be associated with relevant titles ("R").

In order to take into account the whole eye movement trajectory during a task, we implemented a two-stage discriminative HMM, where the first level models transitions between titles, and the second level models transitions between words within a title (topology shown in Fig. 4). Viterbi approximation [20] is used to find the most likely path through the second level model (cf. [21, 22] for similar approaches), and then discriminative training using Extended Baum-Welch is applied to optimize the full model.

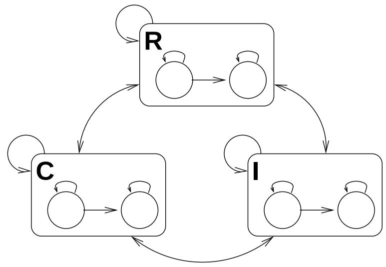

Fig. 4. The topology of the discriminative chain of hidden Markov models.

In our implementation, the first level Markov model has three states, each state corresponding to one class of titles. Each of the three states in the first level have the following exponential family distributions:

1. A multinomial distribution emitting the relevance of the line, B. The parameters of this distribution were fixed, resulting to a discriminative Markov chain model in which each state corresponds to a known classification.

2. A *Viterbi distribution* emitting the probability of the sequence of words in a title.

The Viterbi distribution is defined by the probability of a Viterbi path trough a two-state Markov model forming the second level in our model. The two states of the second level model emit the three exponential observation distributions of Sect. 3.2. The second level Viterbi distributions are further parametrized by the probabilities of beginning the sequence from that state (for example $\Pi^R = \pi_1^R, \pi_2^R$), and transition probabilities between states (e.g. $a_{ij}^R, i, j = 1, 2$). We call the second level Markov model a Viterbi distribution because when evaluating the emission probability we will only take into account the most likely path over the two-state model, called the Viterbi path. After fixing the path the resulting Viterbi distribution is (a fairly complex) exponential family distribution that can be trained with the EBW algorithm. Note that due to the approximation, parameters are optimized only at the states belonging to the Viterbi paths in the second level.

5. DATA ANALYSIS

The simplest method to analyze the eye movement data is to disregard the time dependency between data samples and compute averages of the eye movement features, thus obtaining title-specific feature vectors.

The simplest model using averaged vectors was Linear Discriminant Analysis (LDA), a linear classifier. LDA provided a baseline for more advanced methods such as Support Vector Machine (SVM) [23], which was applied in order to find out how well we can do using only averaged features[2]. The classification results of SVM (see Table 1) show that relevance can indeed be predicted. The ultimate baseline is given by a dumb classifier which assigns all titles to the most likely class.

The deficiency of the simple SVM is that it cannot take the time series nature of the data into account. In order to find more fine-grained cues of relevance (in individual words), we applied HMMs to the data (different models discussed in Section 4).

Classification was performed with leave-one-out validation; each of the task assignments was left for testing in turn, and data from the other tasks was used in teaching. Classification results of different methods are reported in Table 1. Compared to LDA, time series modeling (using HMMs) does improve the results. The data is therefore not too noisy for sequence modelling. It also seems that even with the heuristic training, a global HMM that models the whole trajectory improves the results. This implies that by finding a suitable HMM structure we may also model user behavior. Finally, it is also evident that discriminative modeling improves the classification.

Compared to HMM results, the SVM seems to perform remarkably well, using only averaged features. One reason for this is that SVM is a state-of-the-art method developed explicitly for classification, whereas HMMs form a generative model of all the data, which is of course a different task. Another problem is that large HMMs are prone to overfit the data (even though we used simple models, the amount of data was quite small for probabilistic modeling).

The noise level in the data may be quite high, and its effects on the models is hard to predict. The heuristic assignment of fixations to the closest word may have caused unwanted (noise)effects in segmenting the eye movement trajectories into title-specific sequences. We expect that with a larger data set and better feature extraction the performance of the HMMs is improved.

[2] software package available from http://www.esat.kuleuven.ac.be/sista/lssvmlab/.

Table 1. Predicting relevancy from eye movements. Classification accuracies of leave-one-out validation. Significant (p-value$<$0.01) difference of separate HMMs against the dumb classifier was measured, as well as a significant difference of global HMM against separate HMMs (McNemar's test).

Model	**Accuracy (%)**
Dumb classifier	63.2
LDA	69.2
SVM	75.0
Separate HMMs	71.3
Global HMM	75.8
Discriminative HMM	76.4

6. DISCUSSION

A feasibility study of using time series modeling methods to predict relevancy from eye movements measured in a controlled information retrieval setup was carried out. The first experiments are promising in that even the simple HMM structures with partly heuristic training procedures clearly improve on the results of simple non-sequence models. 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 discover cues about relevance.

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. [24]) and combined with models of document textual content.

Acknowledgments

The work was carried out together with Center for Knowledge and Innovation Research of Helsinki School of Economics, and supported by Academy of Finland, decision 79017. Thanks to Jaana Simola, Ilpo Kojo, Janne Sinkkonen, Nelli Salminen, and Ilkka Kudjoi.[3]

7. REFERENCES

[1] David Tennenhouse, "Proactive computing," *Commun. ACM*, vol. 43, no. 5, pp. 43–50, 2000.

[3] Part of the work has benefited from the Pascal NoE. Access rights to the data sets and other materials are however denied because of other commitments.

[2] Diane Kelly and Jaime Teevan, “Implicit feedback for inferring user preference: a bibliography,” *SIGIR Forum*, vol. 37, no. 2, pp. 18–28, 2003.

[3] Keith Rayner, “Eye movements in reading and information processing: 20 years of research,” *Psychological Bulletin*, vol. 124, no. 3, pp. 372–422, 1998.

[4] Jarkko Salojärvi, Ilpo Kojo, Jaana Simola, and Samuel Kaski, “Can relevance be inferred from eye movements in information retrieval?,” in *Proceedings of the Workshop on Self-Organizing Maps (WSOM'03)*, Hibikino, Kitakyushu, Japan, September 2003, pp. 261–266.

[5] Marcel Adam Just and Patricia A. Carpenter, “The intensity dimension of thought: Pupillometric indices of sentence processing,” *Canadian Journal of Experimental Psychology*, vol. 47, no. 2, pp. 310–339, 1993.

[6] David J. Ward and David J.C. MacKay, “Fast hands-free writing by gaze direction,” *Nature*, vol. 418, pp. 838, 2002.

[7] India Starker and Richard A. Bolt, “A gaze-responsive self-disclosing display,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1990, pp. 3–10, ACM Press.

[8] Aulikki Hyrskykari, Päivi Majaranta, and Kari-Jouko Räihä, “Proactive response to eye movements,” in *INTERACT'03*, G. W. M. Rauterberg, M. Menozzi, and J. Wesson, Eds. IOS press, 2003.

[9] Paul P. Maglio, Rob Barrett, Christopher S. Campbell, and Ted Selker, “Suitor: an attentive information system,” in *Proceedings of the 5th international conference on Intelligent user interfaces*. 2000, pp. 169–176, ACM Press.

[10] Paul P. Maglio and Christopher S. Campbell, “Attentive agents,” *Commun. ACM*, vol. 46, no. 3, pp. 47–51, 2003.

[11] Christopher Campbell and Paul Maglio, “A robust algorithm for reading detection,” in *Workshop on Perceptive User Interfaces (PUI '01)*. November 2001, ACM Digital Library, ISBN 1-58113-448-7.

[12] Manuel G. Calvo and Enrique Meseguer, “Eye movements and processing stages in reading: Relative contribution of visual, lexical and contextual factors,” *The Spanish Journal of Psychology*, vol. 5, no. 1, pp. 66–77, 2002.

[13] Chen Yu and Dana H. Ballard, “A multimodal learning interface for grounding spoken language in sensory perceptions,” in *Proc. ICMI'03*. ACM, 2003, To appear.

[14] Dario D. Salvucci and John R. Anderson, “Automated eye-movement protocol analysis,” *Human-Computer Interaction*, vol. 16, pp. 39–86, 2001.

[15] Miwa Hayashi, “Hidden markov models to identify pilot instrument scanning and attention patterns,” in *Proc. IEEE Int. Conf. Systems, Man, and Cybernetics*, 2003, pp. 2889–2896.

[16] Leonard E. Baum, Ted Petrie, George Soules, and Norman Weiss, “A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains,” *The Annals of Mathematical Statistics*, vol. 41, no. 1, pp. 164–171, February 1970.

[17] D. Povey, P.C. Woodland, and M.J.F. Gales, “Discriminative map for acoustic model adaptation,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP'03)*, 2003, vol. 1, pp. 312–315.

[18] Ralf Schlüter and Wolfgang Macherey, “Comparison of discriminative training criteria,” in *Proc. ICASSP'98*, pp. 493–496. 1998.

[19] Jarkko Salojärvi, Kai Puolamäki, and Samuel Kaski, “Relevance feedback from eye movements for proactive information retrieval,” Tech. Rep. A73, Helsinki University of Technology, Publications in Computer and Information Science, Espoo, Finland, 2003.

[20] A. Viterbi, “Error bounds for convolutional codes and an asymptotically optimum decoding algorithm,” *EEE Transactions on Information Theory*, vol. 13, no. 2, pp. 260– 269, April 1967.

[21] Mikko Kurimo, *Using Self-Organizing Maps and Learning Vector Quantization for Mixture Density Hidden Markov Models*, Ph.D. thesis, Helsinki University of Technology, Espoo, Finland, 1997.

[22] A. Stolcke and S. Omohundro, “Hidden markov model induction by bayesian model merging,” in *Advances in Neural Information Processing Systems 5*, S.J. Hanson, J.D. Cowan, and C.L. Giles, Eds., San Francisco, CA, 1993, pp. 11–18, Morgan Kaufmann.

[23] J.A.K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Squares Support Vector Machines*, World Scientific, Singapore(ISBN 981-238-151-1), 2002.

[24] Ralf Engbert, André Longtin, and Reinhold Kliegl, “A dynamical model of saccade generation in reading based on spatially distributed lexical processing,” *Vision Research*, vol. 42, pp. 621–636, 2002.