Chapter 16

Other projects

16.1 PRIMA—Proactive information retrieval by adaptive models of users' attention and interests

Samuel Kaski, Jarkko Salojärvi, Eerika Savia, Kai Puolamäki

Introduction

Successful proactivity, i.e. 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 PRIMA 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.

In information retrieval 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 the users' intentions. Inference is supported by models of document collections and interest patterns of users.

PRIMA is a consortium with Complex Systems Computation Group, Helsinki Institute for Information Technology (Prof. Petri Myllymäki), and Center for Knowledge and Innovation Research (CKIR), Helsinki School of Economics (Doc. Ilpo Kojo). It started in 2003, and the first results are on modeling of eye movements.

Predicting relevance from eye movements

We measure eye movements during reading, and based on this implicit feedback, try to infer how relevant the document is to the user. Eye movements have earlier been used as alternative input devices in human-computer interfaces (e.g. [5]), and recently in a proactive dictionary which becomes automatically activated [1]. To our knowledge, they have not been used in information retrieval before.

The main challenges are that (i) the signal is complex and very noisy, and (ii) interestingness or relevance is higly subjective and thus hard to define. We started the project by feasibility studies to find out whether the problems are solvable.

We constructed a controlled experimental setting in which it is known which documents are relevant, and then tried to learn relevance from measured eye movement patterns. The user was instructed to find an answer to a specific question, and then shown a set of document titles (Fig. 16.1), of which some were known to be relevant.

In the first feasibility study [3] we extracted a set of standard features [2] from eye movements for each word and combined them to title-specific feature vectors. The two goals of analysing the data were to find out whether relevance can be estimated in this simplified setup using standard features, and which features were important in predicting the relevance. The data was explored with unsupervised methods (Principal Component Analysis and Self-Organizing Maps), and their supervised versions, Linear Discriminant Analysis (LDA) and SOM that learns metrics (cf. Section on Learning metrics).

The results were encouraging; even a simple linear classifier was able to determine relevance clearly better than by chance (80.5% vs. 63%), and a subset of five features was sufficient. There were also many non-linear effects in the data, implicating that a better discrimination is to be expected with a non-linear classifier.

Classification accuracy is also likely to improve when the temporal structure of the data is taken into account. We have started work on Hidden Markov Models (HMMs), which

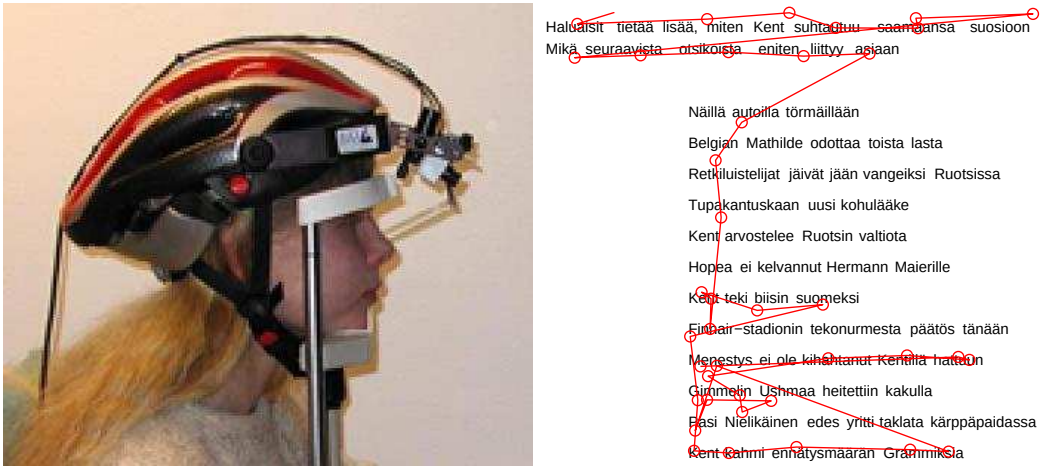

Figure 16.1: The experimental setup. Left: The eye movements of the user are being tracked with a head-mounted eye tracker. The tracker consists of a helmet with two cameras; one monitors the eye and the other one the visual field of the subject. Right: The eye movement pattern during reading plotted on the assignment. Lines connect successive fixations, denoted by circles (Matlab reconstruction). Each line contains one document title, and some of the titles are known to be relevant.

have earlier 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]. First results of applying HMMs to our problem setting show improvement of the classification accuracy from 69.2% (using LDA) to 75.8% (NIPS Machine Learning Meets the User Interface workshop, December 2003).

References

[1] Aulikki Hyrskykari, Päivi Majaranta, and Kari-Jouko Räihä. Proactive response to eye movements. In *Proc. INTERACT'03*. 2003. To appear.

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

[3] 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)*, pages 261–266, Hibikino, Kitakyushu, Japan, September 2003.

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

[5] David J. Ward and David J.C. MacKay. Fast hands-free writing by gaze direction. *Nature*, 418:838, 2002.

[6] 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.

16.2 Data analysis using a tree-shaped neural network

Jussi Pakkanen

Modern data analysis problems usually have to deal with very large databases. When the amount of data samples grow to millions or tens of millions, many traditional tools and techniques slow down noticeably. This, combined with the curse of dimensionality, makes problems involving large data sets very difficult to approach.

Classical computer science has a long history of dealing with data sets. One of the most common approaches is the *divide and conquer* approach, where a large problem is separated into smaller subproblems. Another way to approach the problem is using different kinds of *search trees*, which efficiently index the data.

Our research has focused on finding novel methods to combine neural network systems with large data set manipulation tools of computer science. The goal is to create new neural systems that can be used to analyze huge data bases efficiently while retaining a high precision. The first realization of this research is *The Evolving Tree* [1].

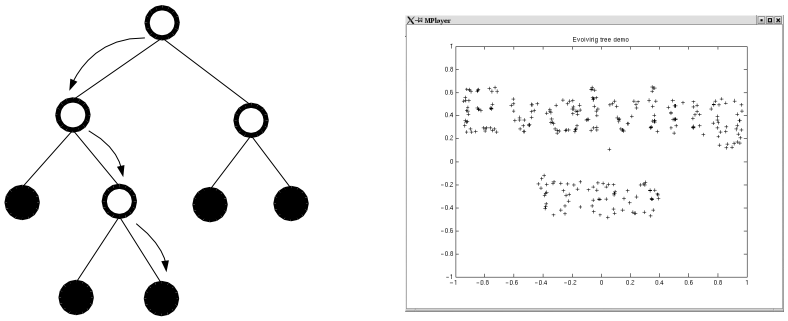

Figure 16.2: The general architecture of the Evolving Tree and an example of adaptation to data.

Figure 16.2 demonstrates the basic properties of the Evolving Tree. The left image shows how the tree is made of two kinds of nodes. The black *leaf nodes* are the actual data analysis nodes, which perform vector coding. The white *trunk nodes* form an efficient search tree to the leaf nodes. The arrows show how a single search on the tree might progress. During training the Evolving Tree grows by creating new leaf nodes to those areas of the data space that are deemed to be underrepresented.

The right image on Figure 16.2 shows how the Evolving Tree adapts to an artificial two-dimensional data set. The dots are the code vectors.. The training had started with a single node, but the tree has grown in size to better explain the data.

Tests on artificial and real world data indicate that the Evolving Tree could be applied to several problems, such as pattern recognition, data mining, density estimation, and exploratory data analysis.

References

[1] J. Pakkanen The Evolving Tree, a new kind of self-organizing neural network, in *Proceedings of the workshop on Self-Organizing Maps '03*, pages 311–316, September 2003, Kitakyushu, Japan.

16.3 Computational model of visual attention

Teuvo Kohonen

By means of simple modeling approaches, an explicit explanation has been given in this work to the following phenomena: 1. Automatic *activation* of a subset of visual signal paths, equivalent to an "attentional window," such that the width of the window is defined by the relative variances of the visual signals. 2. *Narrowing* of the "attentional window" when small saccadic eye movements, voluntary or involuntary, are made. This effect can be shown to ensue from the same model, when the primary signals are further high-pass filtered. 3. *Shifting* of the "attentional window" when strong or novel stimuli (distractors) occur eccentrically in the visual field.

The channel organization

Consider the circular subareas in Fig. 16.3, which delineates a simplified model retina. In this kind of mapping the small foveal areas and the large peripheral areas are thought to project into areas of equal size in the higher parts of the brain. Then we may imagine that the signal paths starting at the retina and ending up on the visual cortex are organized in spatially ordered, functionally separate *channels* corresponding to the small circles in Fig. 16.3. A channel is here identified with a set of signal paths, the *transmittance* of which is controlled by a common *control circuit*. The transmittances of the channels are assumed to have soft shoulders, e.g., Gaussian.

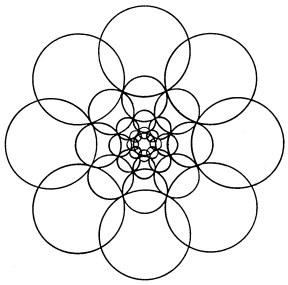

Figure 16.3: Placement of the channels over a hypothetical model retina, around the fovea. A control circuit with corresponding (effective) diameter is associated with each circle.

Assume now that the control circuit of each channel is able to analyze some kind of information content in its incoming signals. The control circuits shall also be able to compare their information contents and mutually *compete* on the permission for activation.

Consider that if we want to compare the information content of subareas relating to such an inhomogeneous sampling system as the retina, any information measure should be related to the resolution of vision in the corresponding subarea.

If the cross section of each channel is then partitioned into an equal number of subfields, if the intensity of the picture is averaged over each such subfield, and if the *variance* of these averaged values is then taken, we obtain a robust measure that is independent of the width of the channel and describes variations of the signal intensity at the given resolution. Let us call this kind of "information measure" the *resolution-related variance*. Notwithstanding, since the absolute variations are slightly different in the light and dark areas of the image, it has further turned out, for photographic images at least, to be most effective to divide the variance by the average of the signal values in the channel.

The sampling grid

In simulations, photographic images were used where the pixels were defined in an orthogonal grid. The resolution-related variance in each channel shown in Fig. 16.3 was computed by placing a *sampling grid* over the channel (Fig. 16.4); the diameter of the sampling grid shall be selected to correspond to the diameter of the due channel, and thus around the assumed direction of the gaze the sampling grid shall be smaller and have fewer pixels, whereas the diameter of the grid shall be selected wider and more pixels must be covered with increasing distance from the direction of the gaze. A constant number, e.g., seven subsets of pixels over each sampling grid were defined, and the averages $av_i, i = 1 \ldots 7$ of the pixels over these subsets were computed. The resolution-related variance for each sampling grid was evaluated by computing the variance of the av_i. After that, the variance was divided by the average of all pixels of this grid. The figure so obtained defines the variable *Variance* in Eq. (16.1).

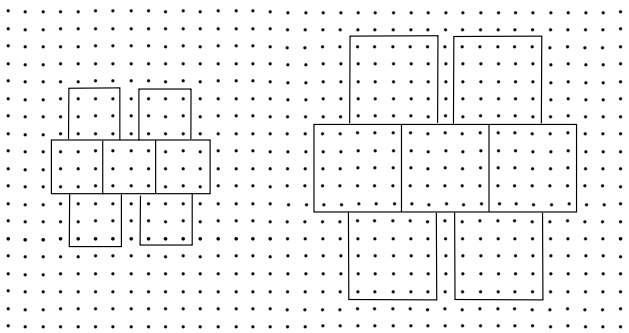

Figure 16.4: Two examples of the sampling grids used for the control of gating of signal transfer in simulations. The small dots correspond to pixels. Over each of the seven square areas, the average $av_i, i = 1 \ldots 7$ of the pixels is computed, whereafter the variance of the av_i is evaluated, and the variance is further divided by the average of pixels over the seven squares.

Optimal width of a channel

Before discussing the system of channels as delineated in Fig. 16.3, it may be interesting to find out how an *optimal width* of a channel, concentrated at a particular location of the image, is determined by the resolution-related variance.

Consider that we try channels of varying width at a certain location of the image. We are looking for the width of the control grid that maximizes the normalized variance of the local averages av_i of the pixel values, denoted *Variance*. Let us call the image data vector **Image**. Let **Grid**(w) mean the choice for the grid with width w; then the "optimal" width w_o is defined to be

$$w_o = \arg\max_w \{ \mathit{Variance}[\mathbf{Grid}(w), \mathbf{Image}] \} . \tag{16.1}$$

A robust optimization of w_o in Eq. (16.1) was carried out over a discrete set of five sampling grids, with their widths varying from 10 to 80 pixels, respectively.

In the first series of simulations illustrated in Fig. 16.5 we demonstrate the "optimal" width of the attentional window, when the gaze was directed at various objects of different widths; the fish, the palm, and the telephone pole, respectively.

Figure 16.5: Demonstration of the opening of attentional windows, the widths of which were automatically determined by the structures present in the area around the gaze. First and third picture: original images. The rest of the pictures show attentional windows, when the gaze was directed to one fish, the palm, and the telephone pole, respectively.

Narrowing of the attentional window

The next phenomenon that is explainable by the optimization approach is the *narrowing of the attentional window* when the gaze is *moved*, voluntarily or involuntarily, by a small amount.

Let us assume that every sampling grid, to some extent, has also high-pass filter properties, i.e., it enhances transient (phasic) values of the signals it samples. Let these temporal variations of the signals ensue from the shifts of the gaze, i.e., translations of the input image over the sampling grids.

Consider the spatial frequencies of the images: if the translation is small, the absolute value of the difference is approximately proportional to the Euclidean norm of its gradient, in which high spatial frequencies are enhanced in proportion to the frequency. In the evaluation of the optimal width w_o from Eq. (16.1), the variances computed from the av_i for the difference image thus decrease with the width of the grid, too, and the optimal width w_o is decreased.

When only a fraction of the previous image is subtracted from the new image, a similar shift of w_o towards smaller values, although a weaker one, can be seen. This effect is then reflected as narrowing of the attentional window. In Fig. 16.6 a sequence of images is

shown, where the subtracted fraction was 50 per cent of each previous image.

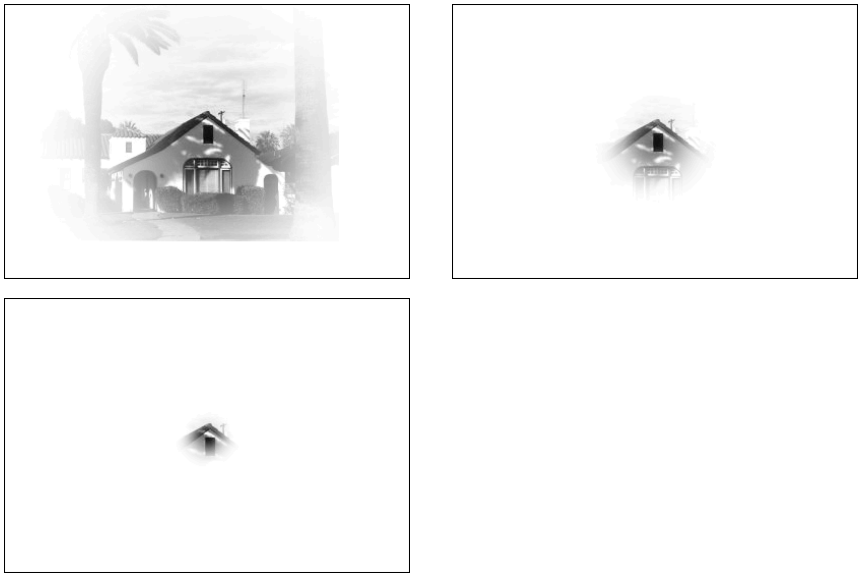

Figure 16.6: Automatic narrowing of the attentional window, when the variances were computed on the basis of images from which 50 per cent of the previously sampled translated image was subtracted. The three pictures form a sequence, in which the gaze was shifted in steps, the size of which became successively smaller.

Attentional window as an activated subset of channels

Finally we shall consider the more complete "biological" case in which the set of channels is fixed and their sizes and positions were defined in Fig. 16.3. For each channel, a sampling grid of corresponding diameter is associated.

Instead of looking for the optimized width of the channel as before, we thus now keep the positions and widths of the channels *fixed* and try to determine a *combination* of k activated channels over which the normalized resolution-related variance of the av_i is highest. In this way, while most of the channels are located eccentrically with respect to the direction of the gaze, the combination of the activated channels defines a more or less symmetric (usually noncircular) attentional window.

In the simulation presented in Fig. 16.7 we thus use the 33-channel "retina" of Fig. 16.3 and let four highest-variance channels define the attentional window. As can be seen, the four channels together tend to emphasize a part of the visual field where some meaningful pattern is present.

It is also discernible that if the variance in the central part of the visual field is low, prominent eccentric patterns tend to *attenuate* weaker parts of the visual field, which can then be interpreted as the *distraction* of attention by the prominent eccentric objects.

References

[1] T. Kohonen, "Modeling of automatic capture and focusing of visual attention," *PNAS*, vol. 99, pp. 9813-9818, 2002.

[2] T. Kohonen, "A computational model of visual attention," Proc. IJCNN'03 (CD-ROM).

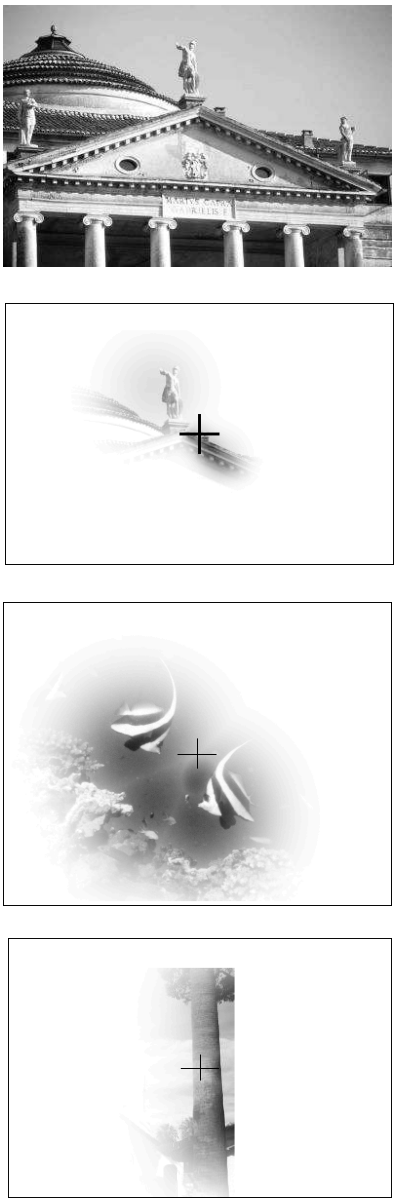

Figure 16.7: Examples of attentional windows spanned by a combination of four activated channels. The black cross indicates the direction of the gaze. The first picture is another original image, of which a part (the statue) is selected and emphasized in the second picture. In the third picture (cf. the first picture in Fig. 16.5, a butterfly-formed attentional window, compassing two of the fishes, is opened. In the lowest picture (cf. the third picture in Fig. 16.5), the form of the resulting attentional window is oblong and the window stretches along the trunk of the tree.