Chapter 9

Proactive Interfaces

Samuel Kaski, Erkki Oja, Jorma Laaksonen, Mikko Kurimo, Arto Klami, Markus Koskela, Mehmet Gönen, Antti Ajanki, He Zhang, Melih Kandemir, Teemu Ruokolainen, Andre Mansikkaniemi, Jing Wu, Chiwei Wang

9.1 Introduction

The Proactive Interfaces research theme combines efforts of multiple research groups, including the Statistical Machine Learning and Bioinformatics group, lead by Professor Samuel Kaski, and the Content-Based Information Retrieval and Speech Recognition groups, lead by Professor Erkki Oja. Since 2008, major collaborative EU FP7, EIT ICT Labs and Aalto funded projects have been carried on which together form the AIRC flagship project *Proactive Interfaces*.

9.2 Inferring interest from implicit signals

Proactive systems anticipate the user's intentions and actions, and utilize the predictions to provide more natural and efficient user interfaces. One of the critical components in this loop is inferring the interests of the user, which is a challenging machine learning problem. Successful proactivity 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.

We have studied inferring interest from eye movement patterns. Eye gaze location is a good proxy for attention but explicit eye movement control is tiresome. Therefore, we study methods that can infer relevance implicitly during normal viewing. Estimated relevance can be used as feedback for an information retrieval system.

We experimented with eye movements and other modalities as source of implicit feedback in image retrieval [1]. It is possible to predict relevant images relatively well from eye movements. We made a feasibility study on predicting the relevance of objects in a video from viewers' eye movements [2]. This setup is an extension of our earlier eye tracking studies on static text and image retrieval setups to dynamic scenes. Even with a relatively simple logistic regression predictor the eye movements predict the relevance with an encouraging accuracy.

The ability to infer relevance in dynamic scenes allows us to do proactive information retrieval in the context of the real world environment [3] which is a novel task. With modern data glasses, which have both augmentation and eye tracking capabilities, it is possible to track the user's attention on real and virtual objects and provide presently relevant information. The data glasses are provided to us by Nokia Research Center (NRC).

Other physiological signals than eye movements are also useful in inferring latent cognitive and emotional state. In [4] we show that learning a combined model of accelerometer, EEG, eye tracker and heart-rate sensors improves prediction accuracy over measurements from individual sensors.

In [5] we introduce a proactive retrieval interface for time-ordered image datasets such as personal lifelogs. Humans can effectively recognize familiar images and use them as reference points when navigating images on a timeline. The system further helps by making relevant images more salient. Relevance is estimated from explicit and implicit mouse movement features.

9.3 Eye-movement enhanced image retrieval

Personal Information Navigator Adapting Through Viewing (PinView)[1] was an EU FP7 funded three-year Collaborative Project coordinated by AIRC. It was started on 1 January 2008 and ended on 31 March 2011. The goal of PinView was a proactive personal information navigator that allows retrieval of multimedia – such as still images, text and video – from unannotated databases. During image browsing and searching with a task-dependent interface, the PinView system infers the goals of the user from explicit and implicit feedback signals and interaction, such as speech, eye movements and pointer traces and clicks, complemented with social filtering. The collected rich multimodal responses from the user are processed with new advanced machine learning methods to infer the implicit topic of the user's interest as well as the sense in which it is interesting in the current context.

The PinView consortium combined pioneering application expertise with a solid machine learning background in content-based information retrieval. Besides AIRC, the project consortium included University of Southampton (UK), University College London (UK), Montanuniversitaet Leoben (AU), Xerox Research Centre Europe (FR), and celum gmbh (AU).

The foremost output of the project was the PinView content-based image retrieval system, that uses (1) the LinRel algorithm for balancing the exploration–exploitation trade-off in image selection, and (2) the Multiple Kernel Learning algorithm for optimal use of available low-level image features for iterative online relevance feedback. The PinView method is able to make use of both explicit relevance feedback, given by pointer clicks on images, and implicit feedback obtained from estimated image relevances based on the user's eye movements while viewing retrieved images. Empirical evaluations have proven the efficiency and scalability of the PinView system in realistic small- and large-scale image retrieval experiments.

In a nutshell, a well-functioning novel search engine was implemented as illustrated in Figure 9.1 and scaling it to huge image collections was found to be feasible. User requirement studies were performed in the initial stage of the project. Later the PinView system was evaluated in four user studies that originated from genuine use case scenarios. These experiments showed that the gaze-based implicit relevance feedback clearly improved image retrieval accuracy and speed compared to the baseline of random browsing. As it can be expected that the price and size of eye tracking devices will continue diminishing while their accuracy and usability are concurrently improving, the effortless combination of browsing and proactive retrieval based on implicit gaze feedback will be useful and available on a large scale. During the project, seven peer-reviewed journal and 32 conference papers have been published by the PinView consortium, including e.g. [1, 6, 7].

9.4 Contextual information interfaces

Contextual information interfaces provide access to information that is relevant in the current context. They use sensory signals, such as gaze patterns, to track the user's context and foci of interest, and to predict what kind of information the user would need at the present time. The information is retrieved from databases and presented in a non-intrusive manner. Main challenges are extraction of context from visual and sensory data,

[1] `http://www.pinview.eu/`

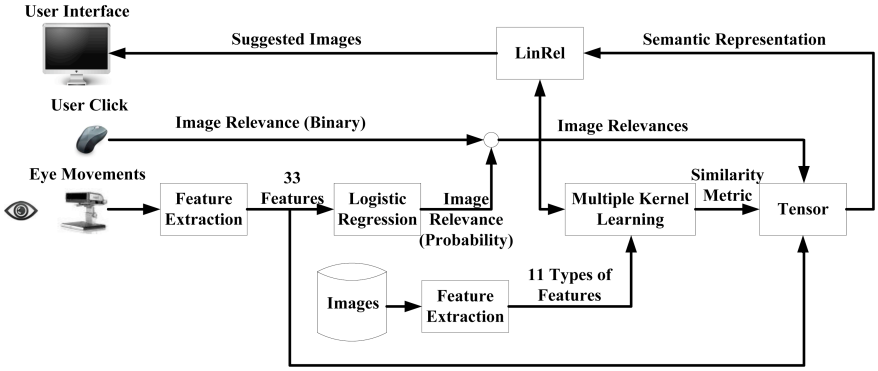

Figure 9.1: Main components and data flow in the PinView content-based image retrieval system that makes use of machine learning of implicit relevance feedback from eye movements.

construction of adaptive machine learning models that are able to utilize heterogeneous context cues to predict relevance, and undisturbing and easily understandable presentation of information. Novel statistical machine learning methods are used for multimodal information retrieval and for taking the context into account.

As a part of *Urban Contextual Information Interfaces with Multimodal Augmented Reality (UI-ART)* project[2], an interdisciplinary research project funded by Aalto *Multidisciplinary Institute of Digitalisation and Energy (MIDE)* programme, we have built a pilot system

[2] `http://mide.aalto.fi/en/UI-ART`

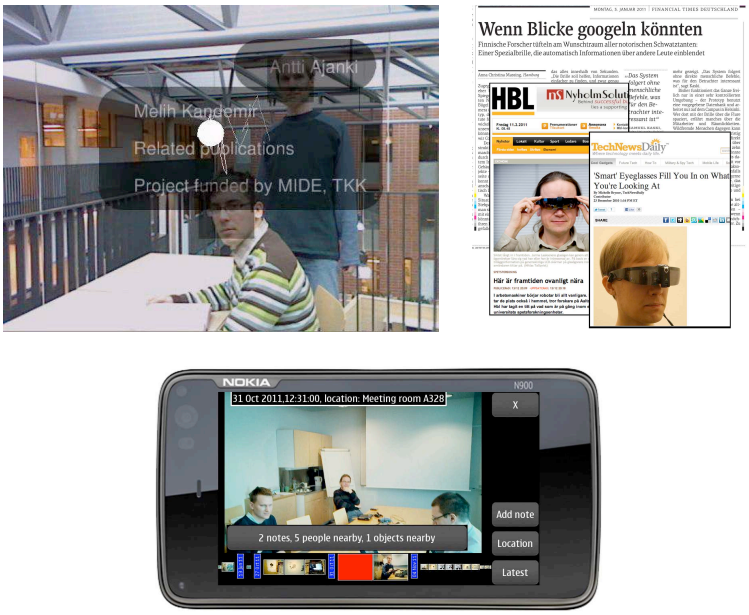

Figure 9.2: Top left: a near-eye display screenshot of the UI-ART contextual information interface. Top right: Our work received international media coverage in 2011. Bottom: Smart phone interface of the UI-ART system.

that retrieves and displays abstract information about people and real world objects in augmented reality [3]. As a pilot application scenario, we have implemented a guide that displays relevant information to a participant in a scientific workshop or meeting or a visitor at a university department. The interface consists of either a head-worn display with an integrated gaze-tracker or a smart phone that can be pointed towards an interesting object. People and objects in the view are recognized from the video feed [8] and information related to them is searched from a database. Retrieved textual annotations are augmented to the view and become part of the context the user can attend to. Evidence from gaze measurements and speech recognition is integrated to infer the user's current interests and annotations that match those are displayed. Figure 9.2 shows snapshots of the UI-ART system's augmented reality display.

In addition to the UI-ART project, the *Proactive Interfaces research group* participated in the *Device and Interoperability Ecosystem (DIEM)* research programme of the TIVIT ICT SHOK from July 2008 to December 2011. The project targeted to enable new services and applications for smart environments that comprise of digital devices containing relevant information for different purposes. The project involved Nokia Research Center (NRC) and Technical Research Centre of Finland (VTT) as collaborators. In 2011, the work was expanded to EIT ICT Labs' Smart Spaces thematic Action Line project *Pervasive Information, Interfaces, and Interaction (PI3)*, where co-operation was been carrier out with research groups from all EIT ICT Labs nodes.

References

[1] Peter Auer, Zakria Hussain, Samuel Kaski, Arto Klami, Jussi Kujala, Jorma Laaksonen, Alex P. Leung, Kitsuchart Pasupa, and John Shawe-Taylor. Pinview: Implicit feedback in content-based image retrieval. In Tom Diethe, Nello Cristianini, and John Shawe-Taylor, editors, *Proceedings of Workshop on Applications of Pattern Analysis*, volume 11 of *JMLR Workshop and Conference Proceedings*, pages 51–57, 2010.

[2] Melih Kandemir, Veli-Matti Saarinen, and Samuel Kaski. Inferring object relevance from gaze in dynamic scenes. In *Proceedings of ETRA 2010, ACM Symposium on Eye Tracking Research & Applications, Austin, TX, USA, March 22-24*, pages 105–108, New York, NY, 2010. ACM.

[3] Antti Ajanki, Mark Billinghurst, Hannes Gamper, Toni Järvenpää, Melih Kandemir, Samuel Kaski, Markus Koskela, Mikko Kurimo, Jorma Laaksonen, Kai Puolamäki, Teemu Ruokolainen, and Timo Tossavainen. An augmented reality interface to contextual information. *Virtual Reality*, 15(2-3):161–173, 2011.

[4] Mehmet Gönen, Melih Kandemir, and Samuel Kaski. Multitask learning using regularized multiple kernel learning. In Bao-Liang Lu, Liqing Zhang, and James Kwok, editors, *Proceedings of 18th International Conference on Neural Information Processing (ICONIP)*, volume 7063 of *Lecture Notes in Computer Science*, pages 500–509, Berlin / Heidelberg, 2011. Springer.

[5] Antti Ajanki and Samuel Kaski. Probabilistic proactive timeline browser. In Timo Honkela, Wlodzislaw Duch, Mark A. Girolami, and Samuel Kaski, editors, *Proceedings of the 21st International Conference on Artificial Neural Networks (ICANN), Part II*, Lecture Notes in Computer Science, pages 357–364, Berlin, 2011. Springer.

[6] Arto Klami. Inferring task-relevant image regions from gaze data. In Samuel Kaski, David J. Miller, Erkki Oja, and Antti Honkela, editors, *Proceedings of IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 101–106. IEEE, 2010.

[7] He Zhang, Teemu Ruokolainen, Jorma Laaksonen, Christina Hochleitner, and Rudolf Traunmüller. Gaze- and speech-enhanced content-based image retrieval in image tagging. In *Proceedings of 21st International Conference on Artificial Neural Networks (ICANN 2011)*, Espoo, Finland, 2011.

[8] Jing Wu. Online face recognition with application to proactive augmented reality. Master's thesis, Aalto University School of Science and Technology, Department of Information and Computer Science, May 2010.