Chapter 9

Proactive Interfaces

Samuel Kaski, Jorma Laaksonen, Mikko Kurimo, Arto Klami, Markus Koskela, Kai Puolamäki, Jarkko Salojärvi, Antti Ajanki, Mats Sjöberg, Ville Viitaniemi, He Zhang, Melih Kandemir, Laszlo Kozma, Lu Wei, Teemu Ruokolainen, Xi Chen, Erkki Oja

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, the Content-Based Information Retrieval group, lead by Docent Jorma Laaksonen, and the Speech Recognition group, lead by Docent Mikko Kurimo. In 2008, three major collaborative projects, PinView, UI-ART and Diem/MMR, have been launched which together form the AIRC flagship project *Proactive Interfaces*. In 2009, a fourth project, Image Based Linking, has been started.

9.2 Inferring interest from gaze patterns

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 focus on inferring the interest and needs of the user from gaze patterns, measured with modern eye-tracking equipment. During complex tasks, such as reading, attention approximately lies on the location of the reader's gaze. Therefore eye movements should contain information, although very noisy, on the reader's interests. As a practical example of what can be inferred from eye movements, [1] uses discriminative Hidden Markov models to detect different processing states in the tasks of simple word search, question-answer, and finding the most interesting topic. The model detects, for example, switches between reading and scanning the text, which in turn helps in predicting the intention of the user.

Another line of work focuses on information retrieval tasks, where the tasks range from estimating relevance of specific text snippets to inferring implicit queries even the user cannot formulate accurately. The eye-movements collected while the user browses the retrieval results are informative of what the user was after, giving an estimated query that can be used for retrieving more relevant documents [2, 3]. The difficult learning problem, termed *learning to learn*, is in finding a regressor from word-level features to queries so that it generalizes to new queries and user interests. [2] solves this by incorporating both the inference of the implicit query and prediction of the relevance of unseen documents into a unified probabilistic model, while [3] utilizes SVM-classifiers in learning the relationships between how words are viewed and their importance for the task. The parameters of the models are optimized to maximize the average performance over a range of training queries, and the resulting query-independent predictors can be applied for topics with no training data.

Going beyond text retrieval tasks, [4] extends the information retrieval work for images, using eye movements for predicting relevance feedback to be used in content-based retrieval. The work is the first demonstration on gaze providing useful information also for media types that are less-structured than text, continued with improved inference and interface in [5].

9.3 Eye-movement enhanced image retrieval

PinView is an EU FP7 funded three-year Collaborative Project started on 1 January 2008 and coordinated by in AIRC. The goal of PinView is 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 will infer the goals of the user from explicit and implicit feedback signals and interaction (eye movements, pointer traces and clicks, speech) 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 combines pioneering application expertise with a solid machine learning background in content-based information retrieval. Besides AIRC, the project consortium includes University of Southampton (uk), University College London (uk), Montanuniversitaet Leoben (au), Xerox Research Centre Europe (fr), and celum gmbh (au). The publications of the PinView project's first two years are [4, 6, 7, 8, 9, 10, 11, 5].

As part of the project, we have developed novel gaze-based interfaces for image retrieval. The purpose is both to create interfaces that can be used without explicit control devices, which is useful for mobile environments and also for people with motor disabilities, but also to obtain more information from gaze. While gaze is informative of the user's interests in all settings, it is possible to create interfaces that provide more information compared to standard displays. We have developed GaZIR [5], a gaze-based zoomable interface for image retrieval, that can be operated with gaze alone. As explicit control with gaze is highly stressful, the GaZIR system uses gaze primarily for implicit information. Explicit control is used only for zooming in and out, while the actual retrieval feedback is learned implicitly from the gaze patterns and fed to the PinView content-based retrieval engine. Figure 9.1 shows a screenshot of the interface, showing the co-centric circles of images the user is currently browsing. The circle-shaped layout was chosen to break natural habits of browsing grid of images in a structured fashion, and hence to extract more information from the gaze trajectory.

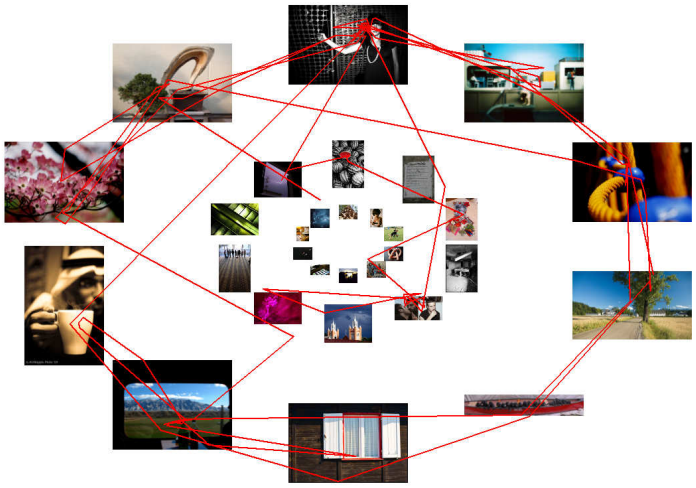

Figure 9.1: GaZIR, the gaze-based zoomable interface for image retrieval, in action. The user sees co-centric circles of images, and the system monitors the gaze of the user (red lines) while he is browsing the images. The subjective relevance of the images is inferred from the gaze trajectory and forwarded to the PicSOM content-based image retrieval system. When the user zooms in, PicSOM returns new sets of images closer to the implicitly defined intents of the user.

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 non-intrusive manner. Main challenges are extraction of context from visual and sensory data, 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, an interdisciplinary research project funded by TKK MIDE (Multidisciplinary Institute of Digitalisation and Energy) programme, we have build a pilot system that retrieves and displays abstract information about people and real world objects in augmented reality [12]. As a pilot application scenario, we have implemented a guide that displays relevant information to a visitor in a university department. The interface consists of either a head-worn display with an integrated gaze-tracker or a hand-held PC that can be pointed towards an interesting object. People and objects in the view are recognized from the video feed 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 a snapshot of the UI-ART system's augmented reality display.

We studied one component of the pilot system, namely prediction of relevance from gaze patterns, in more detail in [13]. We trained a model to predict importance of objects in the scenes of a video, as reported by test subjects, based on gaze patterns recorded

Figure 9.2: The augmented reality of the UI-ART system in the Virtual Laboratory Guide pilot application.

while the subjects were watching the video. In this feasibility study we observed that gaze patterns provide useful information in inferring user interest.

If available, behavior of other people on the same or similar task can be an effective contextual cue. In [14] we introduced a collaborative filtering method that learns a latent structure both for users and documents. With this two-way generalization the model is able to make predictions when either new users or new documents are added to the dataset, unlike earlier state-of-the-art methods.

The Proactive Interfaces research group participates in the Device and Interoperability Ecosystem (DIEM) research programme of the TIVIT ICT SHOK. The project started in July 2008 and targets to enable new services and applications that are based on smart environments that comprise of digital devices containing relevant information for different purposes. The key is interoperability between devices from different domains. Our group is involved in the Mobile Mixed Reality (DIEM/MMR) work package together with the TKK Department of Media Technology, Nokia Research Center (NRC) and Technical Research Centre of Finland (VTT), among others.

The Image Based Linking project began in 2009. The project aims to provide new ways to get access to digital services for mobile phones with integrated digital cameras. This kind of methods can be used for various purposes linking digital information to the physical world. Possible application areas include outdoor advertising, magazine and newspaper advertising, tourist applications, and shopping. In the context of the Proactive Interfaces project, the researched technologies enable more sophisticated object and location recognition for the developed augmented reality applications.

References

[1] Jaana Simola, Jarkko Salojärvi, and Ilpo Kojo. Using hidden markov models to uncover processing states from eye movements in information search tasks. *Cognitive Systems Research*, 9:237–251, 2008.

[2] Kai Puolamäki, Antti Ajanki, and Samuel Kaski. Learning to learn implicit queries from gaze patterns. In Andrew McCallum and Sam Roweis, editors, *Proceedings of ICML 2008, Twenty-Fifth International Conference on Machine Learning*, pages 760–767, Madison, 2008.

[3] Antti Ajanki, David R. Hardoon, Samuel Kaski, Kai Puolamäki, and John Shawe-Taylor. Can eyes reveal interest?—Implicit queries from gaze patterns. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research*, 19:307–339, 2009.

[4] Arto Klami, Craig Saunders, Teófilo de Campos, and Samuel Kaski. *Can relevance of images be inferred from eye movements?*, pages 134–140. ACM, New York, 2008.

[5] László Kozma, Arto Klami, and Samuel Kaski. GaZIR: Gaze-based zooming interface for image retrieval. In *Proc. ICMI-MLMI 2009, The Eleventh International Conference on Multimodal Interfaces and The Sixth Workshop on Machine Learning for Multimodal Interaction*, pages 305–312, New York, NY, USA, 2009. ACM.

[6] He Zhang, Markus Koskela, and Jorma Laaksonen. Report on forms of enriched relevance feedback. Technical Report TKK-ICS-R10, Helsinki University of Technology, Department of Information and Computer Science, Espoo, Finland, November 2008.

[7] Ville Viitaniemi and Jorma Laaksonen. Evaluation of pointer click relevance feedback in PicSOM. Technical Report TKK-ICS-R11, Helsinki University of Technology, Department of Information and Computer Science, Espoo, Finland, November 2008.

[8] Markus Koskela and Jorma Laaksonen. Specification of information interfaces in PinView. Technical Report TKK-ICS-R12, Helsinki University of Technology, Department of Information and Computer Science, Espoo, Finland, November 2008.

[9] Jorma Laaksonen. Definition of enriched relevance feedback in PicSOM. Technical Report TKK-ICS-R13, Helsinki University of Technology, Department of Information and Computer Science, Espoo, Finland, November 2008.

[10] Mats Sjöberg and Jorma Laaksonen. Optimal combination of SOM search in best-matching units and map neighborhood. In *Proceedings of 7th International Workshop on Self-Organizing Maps (WSOM 2009)*, volume 5629 of *Lecture Notes in Computer Science*, pages 281–289, St. Augustine, Florida, USA, 2009. Springer. Available online at: `http://dx.doi.org/10.1007/978-3-642-02397-2\_32`.

[11] Ville Viitaniemi and Jorma Laaksonen. Spatial extensions to bag of visual words. In *Proceedings of ACM International Conference on Image and Video Retrieval (CIVR 2009)*, Fira, Greece, July 2009.

[12] Antti Ajanki, Mark Billinghurst, Melih Kandemir, Samuel Kaski, Markus Koskela, Mikko Kurimo, Jorma Laaksonen, Kai Puolamäki, and Timo Tossavainen. Ubiquitous contextual information access with proactive retrieval and augmentation. Technical Report TKK-ICS-R27, Helsinki University of Technology, Department of Information and Computer Science, Espoo, Finland, December 2009.

[13] Melih Kandemir, Veli-Matti Saarinen, and Samuel Kaski. Inferring object relevance from gaze in dynamic scenes. In *Proc. ETRA 2010, Eye Tracking Research & Applications*, to appear.

[14] Eerika Savia, Kai Puolamäki, and Samuel Kaski. Latent grouping models for user preference prediction. *Machine Learning*, 74:75–109, 2009. Published online: 3 September 2008.