AIR 2006
First International Workshop on Adaptive Information Retrieval

Hideo Joho, Jana Urban, Robert Villa,
Joemon M. Jose, and C. J. van Rijsbergen
Department of Computing Science
University of Glasgow
{hideo,jana,villar,jj,keith}@dcs.gla.ac.uk

http://www.dcs.gla.ac.uk/workshops/air/

Abstract

Adaptive Information Retrieval (IR) systems are designed to optimize retrieval effectiveness and user interaction in an underlying search environment. In this article, we report on a workshop held at Glasgow, UK, in October 2006. The workshop program consisted of three invited talks, four provocative presentations, and a panel discussion. In addition, 15 poster presentations enriched the workshop themes. Important pointers to further development of adaptive IR systems and their evaluation were suggested.

1 Introduction

Information retrieval is a challenging process. Increasingly information is sought from collections with multiple media and genres using multiple modalities and languages. A number of factors affect the retrieval process such as query formulation, matching process and result presentation. However, the importance of context in the retrieval process has been identified. Adaptive retrieval, a process in which the search is adapted towards the user needs/context, has therefore become increasingly important.

The first international workshop on adaptive Information Retrieval (IR) was held at Glasgow, UK, on 16th October 2006. The main objective of the workshop was to bring together researchers in IR and related areas to facilitate the discussion on recent advances and future direction regarding adaptive IR research. The workshop focused especially on researchers who had interest in context sensitive retrieval, personalization systems, adaptive retrieval models, implicit feedback systems, test collection building for interactive evaluation, interactive and adaptive multimedia retrieval and evaluation methodologies.

The workshop program consisted of three keynote talks, four provocative presentations, two discussion groups, and a panel session. The posters from participants were presented during the coffee/lunch breaks. The workshop was successfully held with thirty-one active researchers in IR and related fields from around the globe. This report presents the summary and outcomes of the workshop. The presentation materials are available from the workshop website.

2 The workshop

The workshop began with the first keynote presentation, *Getting Personal: Personalization of Support for Interaction with Information*, given by Nick Belkin (Rutgers University, USA). Belkin first introduced the personalisation of IR as a subset of adaptive IR, followed by a good overview of existing work on this area to date. Belkin then discussed five facets involved in personalisation: relevance/usefulness/interest, task, problem state, personal characteristics, personal preferences, and context/situation. The relevant aspects were presented in each facet. Determining significant aspects, finding means for identifying them, offering effective support, and integrating all facets into single frameworks were argued to be the overall goals for personalisation. Belkin highlighted that there were many opportunities for research, especially on the integration of multiple facets.

The second keynote presentation, *A Model of IR Testing and Evaluation: From Laboratory towards User-Involved*, was given by Noriko Kando (NII, Japan). Kando first presented how the activities at NTCIR (NII Test Collection for IR Systems) had extended the evaluation methodology by considering a user's aspects in their design. Graded relevance assessments, simulated user interaction, and opinion analysis task were shown as the examples of such a direction. Kando then proposed a generic framework of IR testing that was designed to address from a laboratory-based testing to a more user-oriented testing.

Ellen Voorhees (NIST, USA) gave the third keynote presentation, *Building Test Collections for Adaptive Information Retrieval: What to Abstract for What cost?* Voorhees first revisited the design policy of the Cranfield test collection model, and argued that it was the carefully calibrated level of abstracted task that allowed researchers to gain control over variables, and thus, it enabled more experimental power at lower cost. Voorhees then discussed the difficulty of gaining experimental power in interactive settings by looking at past interactive tracks at TREC. Voorhees emphasised the importance of determining the truly distinguishing features (the core competency) of adaptive IR and developing a protocol that captured those features, to step towards an adaptive IR test collection.

Between the three keynote talks, there were four provocative presentations. The first provocative presentation, *Simulating* Searcher's *Feedback Quality and Effort in Interactive IR*, was given by Kalervo Järvelin (Tampere University, Finland). Järvelin first discussed the benefits of simulated interaction such as the diversity of system configurations in testing, cost, flexibility of experimental design, and no learning effects to avoid. Then the simulated user parameters from previous relevance feedback (RF) studies were presented, including relevance threshold, browsing window size, and feedback set size. Järvelin demonstrated that the simulated interaction was useful for determining the level of dependency between the user parameters and retrieval effectiveness of several RF techniques.

Joemon Jose (Glasgow University, UK) gave the second provocative presentation, *Issues in the research on Adaptive search systems*. Jose first showed several demonstrator systems developed at Glasgow that were designed to capture a user's evolving information needs using ostensive models. The user modelling components helped systems to offer a flexible browsing of search results, interactive grouping of retrieved objects for personalised RF, and recommendation of relevant objects based on long-term interests. Having these experiences, Jose emphasised the importance of gathering user interaction data to build an infrastructure for the evaluation of adaptive IR systems.

The third provocative presentation, *Employing User Relevance Assessments for Measuring Retrieval Effectiveness*, was given by David Harper (Robert Gordon University, UK). Harper first presented an exercise of determining the relevance of documents where multiple judgements were available (typical in a user study). It was argued that the difference of interpretation occurred in relevance assessments could be considered as a property of topics, and thus, incorporated into the measure of retrieval effectiveness. Harper suggested looking at the level of relevance agreement to determine the diversity (or ambiguity) of topics, effect of tasks in relevance assessments, and to explore the performance of individual users.

Mark Sanderson (Sheffield University, UK) gave the fourth provocative presentation, *Test collections for all*. Sanderson first argued that the community would need many new test collections to test IR algorithms with

varied contexts. It was also argued that test collections were not necessarily a lot of work to build. Multiple approaches were then discussed to illustrate the formation of test collections (mainly regarding relevance assessment) with limited resources. Sanderson suggested considering the cost-effective techniques for test collections with varied contexts and multiple assessments, and addressing the lack of support for future system when it became a real issue.

The workshop program then went on to the group discussion session. Two groups were formed for an hour discussion: A) *Evaluation methodologies for adaptive IR* and B) *Models of adaptive IR*. Both groups reported back a few points to consider in future work of adaptive IR. The main outcomes were as follows.

Group A: Evaluation methodologies for adaptive IR

- What aspects of the Cranfield model should be emulated regardless of different core competency?
  - Same document collection, tasks, protocol can be used by a user study carried out at each site
- Test collection as a repository of interaction data collected and shared by researchers
  - Reduce efforts of getting user data for automatic evaluation
  - Opportunities to get the data of various contexts for user studies
- How can we use search engines' query logs as evaluation resource?
  - How can we complement the basic query and click-through data with contextual information, thus, create a better quality collection?
- Can we use recall/precision or new measures should be devised?

Group B: Models of adaptive IR

- Information needs are dynamic – how can we reduce the variability and ambiguity of queries?
  - Strengthen the dialogue between a user and search engine can be seen as an overall goal of adaptive IR
  - If we see an underlying retrieval model as a given, what interactive support and interface can we provide to facilitate the system's adaptation?
- Can multiple relevance assessments be used to model the adaptiveness of systems?
- Rather focusing on document retrieval, finding related term or similar task might be more appropriate to test whether or not a technique reduced the variability or ambiguity of queries
- Measure of system adaptation
  - Search steps – can the number of steps taken in search indicate the level of adaptation?
  - Control – should a user always have a control of search, or should a system take over the control and do the adaptation under the scene? What is the effect of tasks on the control?

C. J. van Rijsbergen (Glasgow University, UK) chaired a panel discussion as the final part of the workshop program. The panel consisted of Kalervo Järvelin, Mark Sanderson, David Harper, and Nick Belkin. The panellists first presented their view on adaptive IR and evaluation, followed by an open discussion with other participants. The major outcomes, including the responses to the group discussion, were as follows.

- Scope of test collections for adaptive IR
  - Shared resources of interaction data (standard markup scheme for interaction data)
  - Understanding information seeking behaviour (inc. rational and irrational)
  - Understanding significant contextual features that can help systems be adaptive
  - Measuring interaction power of systems/interfaces
- Combination with other resources
  - Worth looking at a complementary use of query logs for adaptive IR testing before spending a lot of effort creating new logs
  - Deploy a search service/tool to gather log data (for subsequent analysis)
- Models of relevance assessments
  - Beyond a single interpretation of relevance in development and testing
  - One or two topics with hundreds of users?
  - Diversity of users (i.e., interpretation) should be maximised

As can be seen, there were certainly more questions than answers. However, as one of the panellists commented, this was a great opportunity for the community to look at these issues more closely and to advance the existing IR technologies.

3 Conclusion

This report presented a summary of the first international workshop on adaptive IR held at Glasgow in October 2006. With the major researchers in IR involved as active participants, many interesting ideas were discussed during the workshop program. Most importantly, it became evident that the area of adaptive IR, while it is still at an early stage, contains many potential research directions that encourage researchers to explore new aspects of information retrieval. Looking at the specification of ideal test collections for adaptive IR was suggested as one example.

Finally, it should be noted that a special topic issue on adaptive IR is being published in the journal of Information Processing and Management where the workshop organisers are involved as the guest editors.

4 Acknowledgements

Special thanks go to the speakers and all participants for contributing to a successful workshop. We wish to thank the UK Multimedia Knowledge Management Network (EPSRC Ref: GR/T02690/01) and Department of Computing Science, University of Glasgow for their financial support of the workshop.