Dual Interactive Information Retrieval

Vitaliy Vitsentiy
Queensland University of Technology
National ICT Australia

v.vitsentiy@qut.edu.au

Abstract *A new task in Interactive Information Retrieval (IIR) is considered – optimization of information retrieval taking into account impact on quality of interaction with the user. Dual IIR is defined.*

Keywords dual interactive information retrieval, multistage stochastic programming.

1 Introduction

Information Retrieval Systems (IRSs) are used for search of indistinct information. The user has some information need, which they translate into some query language that usually can not describe the information need precisely. From the point of view of the IRS there is an uncertainty what is searched by the user. This is caused by: 1) information need is not precise; 2) information need is not represented precise enough by the query; and 3) information need changes. So, IRSs make decisions about which documents to select based on the user query in situation where there is not enough information for unambiguous decisions. IRS usually retrieves not one document but a portion of documents, among which the user may find the necessary document. Often information need is not satisfied at once and the search process goes on (see Figure 1). Thus Interactive Information Retrieval (IIR) may be considered as a problem of multistage decision making under uncertainty. Such problems are usually solved by stochastic programming methods [2].

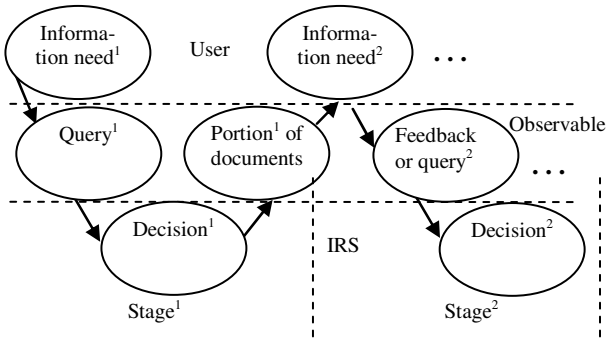

Figure 1: Dynamics of information search.

2 Definition

Dual IIR (DIIR) is information retrieval with feedback after each stage of retrieval when at selection of documents the IRS takes into account not only their relevance but also possible effect of the retrieved portion of documents on user feedback for optimization of the whole user search session. The term "dual" is taken by analogy with "dual control" from adaptive control theory. The dual features are: 1) direction: DIIR System retrieves more relevant documents in the current portion; 2) probing: DIIR System encourages receiving better user feedback. The need for duality is caused: 1) user information need is uncertain and the uncertainty should be reduced; 2) the feedback is evaluative, not instructional and therefore it does not allow determining whether there are more relevant documents among the not retrieved documents than the retrieved documents. The expediency of the development of the methods is based on the hypothesis: if user feedback is encouraged to be better especially on early stages of search, even by retrieving documents with somewhat lower relevance, then IRS using feedback can estimate relevance of documents more precisely and this leads to better relevance of all documents in the whole user search session.

Retrieving documents in DIIR is not simple selection of documents with the largest relevance but is an optimization problem. Because the optimal value is determined by: 1) probing may lead to lower relevance of documents in the current portion but a better feedback will allow to determine relevance better and retrieve more relevant documents on the next stages; 2) an increase of quantity of retrieved similar documents may lead not only to increase of relevance of the current portion but also to decrease of feedback quality and accordingly to decrease of relevance on the next stages.

The problem of DIIR includes into the IR decision making also the task similar to active learning [3]. So, DIIR is a problem of adaptive dual control [1], where the problem of optimal balance between control and estimation is studied. A mathematic programming model for decision making in DIIR can be found in [4].

3 Conclusion

DIIR is defined. A positive effect from usage of DIIR should be expected in the situation of vague user information need, long search interaction process, broad semantic range of search.

References

[1] A.A. Fel'dbaum, "Dual Control Theory. I-IV" *Automation Remote Control*, 21, 22, pp. 874-880, 1033-1039, 1-12, 109-121, 1960-1961.

[2] *Stochastic Programming*, ed. by Ruszczynski A., and Shapiro A., Elsevier, 2003.

[3] S. Tong, *Active learning: theory and applications*, PhD thesis, Stanford University, 2001.

[4] V. Vitsentiy, "A decision making model for dual interactive information retrieval", *International Conference on Systems, Computing Sciences and Software Engineering 2006*, in press.