Conceptualisation of a Belief Revision approach to Information Retrieval

David E. Losada and Alvaro Barreiro
Dept. Computer Science
University of A Corunna
Spain
email:{losada,barreiro}@dc.fi.udc.es

Abstract. In this work we present the conceptualisation of a recent Belief Revision approach to Information Retrieval. After a brief introduction of the approach, the conceptualisation is presented. It follows the framework proposed by Lalmas and Bruza for comparison of different logical approaches to Information Retrieval. Finally, future lines of work are commented.

1 Belief Revision and Information Retrieval

Belief Revision (BR) deals with the issue of updating a knowledge base in the light of new information. The trivial case appears when the new information does not contradict the previous knowledge, and the revision simply expands the knowledge base with the new information. When conflicts arise, the basic policy is that the revised knowledge base must imply the new information and retain as much old knowledge as possible. Therefore, some criterion is needed in order to decide what parts of the new information are *closest* to the old information. *Model-based* approaches to BR establish an order among the models of the new information induced by the knowledge base. Models of the revised theory are the minimal models of the new information in this order. If we regard the representation of a query q as a theory and a document representation d as a new information, the BR process $q \circ d$ produces a measure of the uncertainty of $d \rightarrow q$, in the line proposed by van Rijsbergen in [7]. The application of this BR process to the field of Information Retrieval (IR) has been introduced in a recent work [6]. The BR operator suggested by Dalal [3] , $\circ_D$, was proposed to carry out the revision, $q \circ_D d$. An important result in this direction is that using Dalal's operator ensures theoretically an ordering among the documents which follows the notion of proximity to the query. In [6] the BR framework was used to build a similarity measure, $BRsim$, that was shown equivalent to the inner product matching function, which is frequently used in the vector-space model. However, this equivalence was done restricting the expressiveness of the query. This decision was made in order to get some clear results showing that classical measures can be modeled within a logical framework. Disjunctions in queries were treated as conjunctions, so that the representation was equivalent to a vector of terms. Nevertheless, the framework supplied by BR could be used to build more complex measures using the total expressiveness of the Propositional Logic. In this case and from an IR perspective, the framework would be a Boolean Model extended with measures in the interval [0,1]. In addition, the model can deal with partial representations of documents, which is the case when not every term appears in a document representation. This is not a usual characteristic of IR systems.

2 Conceptualisation

Since van Rijsbergen expressed the necessity of quantifying the uncertainty of the relevance of the document to the query [7], several logical approaches have addressed this task. Recent works [5, 4, 2] present an analysis of these approaches, showing that the question of which logic is the appropriate to quantify the relevance is still subject to research. In [5] a conceptual framework is proposed to compare the different logical models to IR. We think it is interesting to analyze our proposal within this framework in order to understand properly its features.

The conceptual framework is based on the evaluation of the way that different approaches implement the logical uncertainty principle. This principle states that:

> " Given any two sentences x and y; a measure of the uncertainty of $y \to x$ relative to a given data set, is determined by the minimal extent to which we have to add information to the data set, to establish the truth of $y \to x$ "

The conceptual components used are the document representation, d, the query representation, q, the relevance $d \to q$, the data set and the transformation needed to establish the truth of $d \to q$. The implementation of this principle within our proposal produces the conceptual components presented in fig.1.

Document representation d	Propositional formula d conjunction of literals
Query representation q	Propositional formula q conjunction of literals
Relevance $d \to q$	BRsim(d,q)
Data set	None
Transformation	Revision $q \circ_D d$

Fig. 1. Conceptual components of the logical uncertainty principle

The relevance is computed by $BRsim$, which is equivalent to the inner product query-document matching function. Nevertheless, as stated above, the BR framework is open to define another similarity measures. Moreover, query and document representations that have conjunctions and disjunctions can be directly used within the framework.

An important point is that the model is open to be extended by introducing structures like thesauri expressing relationships among terms. These structures would constitute the data set in Lalmas' terminology.

3 Future work

Model-based Belief Revision is intractable in general, due to the exponential growth of the number of models. The search for methods to reduce this complexity is a hot topic in the BR community. In this line of research the work by Chou and Winslett [1] shows that restricted problems can be solved within reasonable bounds. Indeed, they present a system which implements the revision and the expected running time is polynomial for many interesting cases.

We are working now in the development of an algorithm for the BR process proposed for IR. Representations of documents are already restricted to conjunctions so that the complexity is expected to be reduced.

Another line of research is based on increasing the expressiveness of the model. Current representation of terms is reduced to presence or absence. It is not possible to represent levels of presence. Issues like representing term weights in the interval [0,1] need formalisms more expressive than Propositional Logic. First order logic or its sublanguages should be taken into account in this research.

Acknowledgements

This work was supported in part by "Ministerio de Educacion y Cultura" - Spain, project PB97-0228 and by Government of Galicia, project PGIDT99PX110201B.

References

1. T. S-C. Chou and Winslett M. Immortal:a model-based belief revision system. In *Proceedings of KR-91, the Second Conference on Principles of Knowledge Representation and Reasoning*, pages 99–110, Cambridge,Massachusetts, April 1991.
2. F. Crestani, M. Lalmas, and C.J.Van Rijsbergen, editors. *Information Retrieval, Uncertainty and Logics: advanced models for the representation and retrieval of information.* Kluwer Academic Publishers, Norwell, MA, 1998.
3. M. Dalal. Investigations into a theory of knowledge base revision: Preliminary report. In *Proceedings of AAAI-88, the 7th National Conference on Artificial Intelligence*, pages 475–479, 1988.
4. M. Lalmas. Logical models in information retrieval: introduction and overview. *Information Processing & Management*, 34(1):19–33, 1998.
5. M. Lalmas and P. Bruza. The use of logic in information retrieval modeling. *Knowledge Engineering Review*, 13(3):263–295, 1998.
6. D. E. Losada and A. Barreiro. Using a belief revision operator for document ranking in extended boolean models. In *Proceedings of SIGIR-99, the 22th ACM Conference on Research and Development in Information Retrieval*, pages 66–73, Berkeley, California, August 1999.
7. C.J. Van Rijsbergen. A non-classical logic for information retrieval. *The Computer Journal*, 29:481–485, 1986.