Information Retrieval Estimation via Fuzzy Probability

Zhiheng Huang

Department of Computer Science
The Australian National University
Australia

zhiheng@cs.anu.edu.au

Tamás D. Gedeon

Department of Computer Science
The Australian National University
Australia

tom@cs.anu.edu.au

Aim: Fuzzy logic [3] is a useful approach to help identify partially matched documents for a given query. Research on fuzzy information retrieval mainly focuses on the relevance probability estimation of the retrieved documents but lacks the estimation of imprecision of such probability. This study attempts to use fuzzy probability to evaluate the relevance of retrieved documents.

Methods: Probability is widely used in information retrieval to estimate the relevance of retrieved documents. As *relevance* cannot be defined accurately, i.e., it is subjective in practice, the probability estimate of the retrieved documents cannot be precise. This suggests the use of an imprecise representation of probability estimation, which is termed *fuzzy probability*.

One example (as Fig. 1) is to show the fuzzy probability of a document to be "very relevant" to a query. As the definition of "very relevant" is imprecise, the probability of such a retrieved document is hence imprecise. As can be seen, it is *certain* ($possibility = 1$) for such a document to be "very relevant" with a probability of 30%, and it is *quite certain* ($possibility = 0.8$) for it to be "very relevant" with a range of probability from 20% to 35%. The higher possibility value leads to the narrower estimation of probability. In the extreme case, the peak point in the fuzzy probability corresponds to the conventional probability estimate.

There are considerable studies on fuzzy probability. For example, the possibility-probability distribution [1] has been recently proposed. The problem of decision making [2] in the face of a fuzzy probability estimate is investigated.

Results: We have proposed a novel method to calculate fuzzy probability. This method will be applied to calculate a document's relevance (in terms of fuzzy probability) for a query. By using the calculated fuzzy probabilities, different crisp probabilities can be computed [2] with respect to different uncertainty levels (such as *certain* and *quite certain*), which provides a flexible way to evaluate the relevance of a given document.

Conclusions: This study has proposed a novel method to calculate fuzzy probability. At this stage, no experimental work has yet been carried out. The experimental evaluation of this method in information retrieval will be our most important future work.

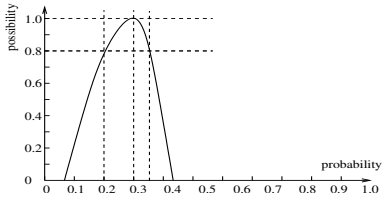

Figure 1: A fuzzy probability estimation of a document to be "very relevant"

References

[1] C. F. Huang and Y. Xue. Some concepts and methods of information granule diffusion. In *IEEE International Conference on Granular Computing*, Volume 1, pages 28 – 33, 2005.

[2] R. R. Yager. Decision making with fuzzy probability assessments. *IEEE Transactions on Fuzzy Systems*, Volume 7, 1999.

[3] L. A. Zadeh. Fuzzy sets. *Information and Control*, Volume 8, 1965.

Proceedings of the 10th Australasian Document Computing Symposium, Sydney, Australia, December 12, 2005.