Similarity Management for Fuzzy Data Mining

Bernadette Bouchon-Meunier

LIP6, Université Pierre et Marie Curie-Paris 6,
104 avenue du Président Kennedy, 75016 Paris, France

Bernadette.Bouchon-Meunier@lip6.fr

Abstract

Data mining is a domain difficult to cope with for various reasons. First, most of the databases are complex, large, and contain heterogeneous, imprecise, vague, uncertain, incomplete data. Furthermore, the queries may be imprecise or subjective in the case of information retrieval, the mining results must be easily understandable by a user in the case of data mining or knowledge discovery. Fuzzy logic provides an interesting tool for such tasks, mainly because of its capability to represent imperfect information, for instance by means of imprecise categories, measures of resemblance or aggregation methods. We will focus our study on the use of similarity measures which are key concepts for many steps of the process, such as clustering, construction of prototypes, utilization of expert or association rules, fuzzy querying, for instance.

We will consider a general framework for measures of comparison, compatible with Tversky's contrast model, providing tools to identify similar or dissimilar descriptions of objects, for instance in a case-based reasoning or a classification approach.

We present some real-world problems where these paradigms have been exploited among others to manage various types of data such as image retrieval or risk analysis.

References

[1]   Bouchon-Meunier, B., Detyniecki, M., Lesot, M.-J., Marsala, C. & Rifqi, M., Real world fuzzy logic applications in data mining and information retrieval, in: *Fuzzy Logic - A Spectrum of Theoretical and Practical Issues.* P. P. Wang, D. Ruan & E. E. Kerre (Eds), (Springer), pp. 219-247, 2007.

[2]   B.†Bouchon-Meunier, M.†Rifqi, S.†Bothorel, Towards general measures of comparison of objects, *Fuzzy Sets and Systems*, 84:(2):143-153, 1996.

[3]   M.-J. Lesot, M. Rifqi, B. Bouchon-Meunier, Fuzzy prototypes: from a cognitive view to a machine learning principle, *in* H. Bustince, F. Herrera, J. Montero (eds.) *Fuzzy Sets and Their Extensions: Representation, Aggregation and Models .Studies*, Springer Verlag, in press

[4]   J. F. Omhover, M. Detyniecki, B. Bouchon-Meunier, A Region-Similarity-Based Image Retrieval System, *Proc. International Conference IPMU'04*, Perugia, Italy, 2004.

[5]   J-F Omhover, M. Rifqi, M. Detyniecki, Ranking Invariance based on Similarity Measures in Document Retrieval, *AMR 2005*, Glasgow, U.K. Springer Lecture Notes LNCS 3877, pp. 55-64, 2006.

[6]   M.†Rifqi, Constructing prototypes from large databases, *Proc. International Conference IPMU'96*, Granada, 301-306, 1996.

[7]   M. Rifqi, V. Berger, B. Bouchon-Meunier, Discrimination power of measures of comparison, *Fuzzy Sets and Systems*, 110(2):189-196, 2000.

[8]   A. Tversky, Features of similarity, *Psycho. Rev.* 84(4):327-352, 1977.

[9]   L. A. Zadeh, Similarity relations and fuzzy ordering, Information Science, 177-200, 1971.