© 2013 FIZ Karlsruhe

io-port 06080918

Liang, Jiye; Wang, Feng; Dang, Chuangyin; Qian, Yuhua

An efficient rough feature selection algorithm with a multi-granulation view.

Int. J. Approx. Reasoning 53, No. 6, 912-926 (2012).

Summary: Feature selection is a challenging problem in many areas such as pattern recognition, machine learning and data mining. Rough set theory, as a valid soft computing tool to analyze various types of data, has been widely applied to select helpful features (also called attribute reduction). In rough set theory, many feature selection algorithms have been developed in the literatures, however, they are very time-consuming when data sets are in a large scale. To overcome this limitation, we propose in this paper an efficient rough feature selection algorithm for large-scale data sets, which is stimulated from multi-granulation. A sub-table of a data set can be considered as a small granularity. Given a large-scale data set, the algorithm first selects different small granularities and then estimate on each small granularity the reduct of the original data set. Fusing all of the estimates on small granularities together, the algorithm can get an approximate reduct. Because of that the total time spent on computing reducts for sub-tables is much less than that for the original large-scale one, the algorithm yields in a much less amount of time a feature subset (the approximate reduct). According to several decision performance measures, experimental results show that the proposed algorithm is feasible and efficient for large-scale data sets.

Keywords: feature selection; multi-granulation view; rough set theory; large-scale data sets
doi:10.1016/j.ijar.2012.02.004