

Logical analysis of data: Estimation of cause-effect relationship under noise with applications in classification of natural water springs

Mariya Naumova, András Prékopa

Rutgers Business School, Rutgers University, Piscataway, New Jersey, United States
mnaumova@business.rutgers.edu

Abstract

Logical analysis of data (LAD) is a methodology introduced by Peter L. Hammer in 1986, aimed at discovering hidden structural information in data sets by using the theory of partially defined Boolean functions. A modification of the approach involving estimation of the cause-effect relationship under noise was proposed by Prékopa in 1994 and turned out to be efficient for data sets for which the sets of positive and negative observations can be separated with a good accuracy by a hyperplane. This paper develops a technique that combines Prékopa's method with initial data clustering. We illustrate the method on water quality monitoring data set in Perm, Russia.