© 2013 FIZ Karlsruhe

io-port 01970954

Ho, Tu Bao; Nguyen, Trong Dung; Shimodaira, Hiroshi; Kimura, Masayuki

A knowledge discovery system with support for model selection and visualization.

Appl. Intell. 19, No. 1-2, 125-141 (2003).

Summary: The process of knowledge discovery in databases consists of several steps that are iterative and interactive. In each application, to go through this process the user has to exploit different algorithms and their settings that usually yield multiple models. Model selection, that is, the selection of appropriate models or algorithms to achieve such models, requires meta-knowledge of algorithm/model and model performance metrics. Therefore, model selection is usually a difficult task for the user. We believe that simplifying the process of model selection for the user is crucial to the success of real-life knowledge discovery activities. As opposed to most related work that aims to automate model selection, in our view model selection is a semiautomatic process, requiring an effective collaboration between the user and the discovery system. For such a collaboration, our solution is to give the user the ability to try various alternatives and to compare competing models quantitatively by performance metrics, and qualitatively by effective visualization. This paper presents our research on model selection and visualization in the development of a knowledge discovery system called D2MS. The paper addresses the motivation of model selection in knowledge discovery and related work, gives an overview of D2MS, and describes its solution to model selection and visualization. It then presents the usefulness of D2MS model selection in two case studies of discovering medical knowledge in hospital data-on meningitis and stomach cancer-using three data mining methods of decision trees, conceptual clustering, and rule induction.

Keywords: KDD; model selection; visualization; user's participation
doi:10.1023/A:1023876925609