

Studies of Rule Induction by STRIM from the Decision Table with Contaminated Attribute Values from Missing Data and Noise – in the Case of Critical Dataset Size –

Authors : Tetsuro Saeki, Yuichi Kato, Shoutarou Mizuno

Abstract : STRIM (Statistical Test Rule Induction Method) has been proposed as a method to effectively induct if-then rules from the decision table which is considered as a sample set obtained from the population of interest. Its usefulness has been confirmed by simulation experiments specifying rules in advance, and by comparison with conventional methods. However, scope for future development remains before STRIM can be applied to the analysis of real-world data sets. The first requirement is to determine the size of the dataset needed for inducting true rules, since finding statistically significant rules is the core of the method. The second is to examine the capacity of rule induction from datasets with contaminated attribute values created by missing data and noise, since real-world datasets usually contain such contaminated data. This paper examines the first problem theoretically, in connection with the rule length. The second problem is then examined in a simulation experiment, utilizing the critical size of dataset derived from the first step. The experimental results show that STRIM is highly robust in the analysis of datasets with contaminated attribute values, and hence is applicable to realworld data.

Keywords : rule induction, decision table, missing data, noise

Conference Title : ICGCRSIS 2015 : International Conference on Granular Computing, Rough Sets and Intelligent Systems

Conference Location : Copenhagen, Denmark

Conference Dates : June 11-12, 2015