Research and Reviews: Journal of Global Research in Computer Science

Extended Abstract
Vol. 9, Iss. 2
2018

Data Mining 2016:Extensions of dynamic programming for decision tree study- Mikhail Moshkov , King Abdullah University of Science and Technology

Mikhail Moshkov
King Abdullah University of Science and Technology, Saudi Arabia

In the presentation, we consider extensions of dynamic programming approach to the study of decision trees as algorithms for problem solving; as a way for knowledge extraction and representation, and as classifiers which for a new object given by values of conditional attributes, define a value of the decision attribute. These extensions allow us: (i) To describe the set of optimal decision trees; (ii) To count the number of these trees; (iii) To make sequential optimization of decision trees relative to different criteria; (iv) To find the set of Pareto optimal points for two criteria; and (v) To describe relationships between two criteria. The results include the minimization of average depth for decision trees sorting eight elements (this question was open since 1968), improvement of upper bounds on the depth of decision trees for diagnosis of 0-1 faults in read-once combinatorial circuits; existence of totally optimal (with minimum depth and minimum number of nodes) decision trees for Boolean functions; study of time-memory tradeoff for decision trees for corner point detection; study of relationships between number and maximum length of decision rules derived from decision trees; study of accuracy-size tradeoff for decision trees which allows us to construct enough small and accurate decision trees for knowledge representation; and decision trees that as classifiers, outperform often decision trees constructed by CART. The end of the presentation is devoted to the introduction to KAUST.

This thesis is devoted to the development of extensions of dynamic programming to the study of decision trees. The considered extensions allow us to make multi-stage optimization of decision trees relative to a sequence of cost functions, to count the number of optimal trees, and to study relationships: cost vs cost and cost vs uncertainty for decision trees by construction of the set of Pareto-optimal points for the corresponding bi-criteria optimization problem. The applications include study of totally optimal (simultaneously optimal relative to a number of cost functions) decision trees for Boolean functions, improvement of bounds on complexity of decision trees for diagnosis of circuits, study of time and memory trade-off for corner point detection, study of decision rules derived from decision trees, creation of new procedure (multi-pruning) for construction of classifiers, and comparison of heuristics for decision tree construction. Part of these extensions (multi-stage optimization) was generalized to well-known combinatorial optimization problems: matrix chain multiplication, binary search trees, global sequence alignment, and optimal paths in directed graphs.

Decision trees are widely used as predictors , as a way of knowledge representation and as algorithms for problem solving . To have more understandable decision trees we need to minimize the number of nodes in a tree. To have faster decision trees we need to minimize the depth or average depth of a tree. In many cases, we need to minimize the number of misclassifications for a tree under some restrictions on time or space complexity of the tree. If we would like to minimize the number of decision rules derived from the tree, we need to minimize the number of terminal nodes in the tree. Unfortunately, almost all problems connected with decision trees optimization are NP-hard.

Biography

Mikhail Moshkov is a Professor in the CEMSE Division at King Abdullah University of Science and Technology, Saudi Arabia. He earned his Master's degree from Nizhni Novgorod State University, received his Doctorate from Saratov State University, and Habilitation from Moscow State University. In 2003, he has worked at the Institute of Computer Science, University of Silesia, in Poland. His main areas of research are Complexity of Algorithms, Combinatorial Optimization, and Machine Learning. He has published 5 research papers in Springer.

Email: mikhail.moshkov@kaust.edu.sa