Documentation: MIP Function — Fitted Rule Ensembles

The rule ensemble algorithm for predicting structured target variables.

Metadata

Category	MIP function, data analysis algorithm
Maintainers	Martin Breskvar
	Bernard Ženko
Homepage	http://source.ijs.si/hbp/clus/wikis/home
Documentation	http://source.ijs.si/hbp/mipfunctions/raw/master/doc/r-clus-rules.pdf
	http://source.ijs.si/hbp/clus/wikis/documentation
Support	http://source.ijs.si/hbp/clus/issues
Source Code	http://source.ijs.si/hbp/clus.git
License	GNU GPL
Current Version	2.12
All Versions	2.12

Description

Methods for learning decision rules are being successfully applied to many problem domains, in particular when understanding and interpretation of the learned model is necessary. In many real life problems, we would like to predict structured target variables, e.g., multiple related numeric variables, or time series. While several methods for learning rules that predict multiple targets at once exist, they are all based on the covering algorithm, which does not work well for regression problems. A better solution for regression is the rule ensemble approach that transcribes an ensemble of decision trees into a large collection of rules. An optimization procedure is then used to select the best (and much smaller) subset of these rules and to determine their respective weights.

This MIP function implements the FIRE algorithm, which employs the rule ensembles approach for solving multi-target regression and time series problems. We can improve the accuracy of the rule model by adding simple linear functions to the ensemble, which results in a model that is a combination of global linear functions and rules.

The development of the algorithm and its implementation was partly paid by HBP.

References

[1] T. Aho, B. Ženko, S. Džeroski, T. Elomaa: Multi-Target Regression with Rule Ensembles; Journal of Machine Learning Research, 13(Aug): 2367–2407, 2012.

[2] B. Ženko, J. Struyf, S. Džeroski: Analyzing time series gene expression data with predictive clustering rules. Machine learning in systems biology: proceedings of the Third International Workshop. University of Helsinki, pp. 177-178, 2009.

Usage

Input: A set of examples with known values for descriptive and target variables.

Parameters: Descriptive variables, target variables, maximum number of rules.

Output: A predictive model written as a combination of a linear equation and decision rules.

Example

We can illustrate the use of the algorithm on the ADNI data.

Task:

Given a data set consisting of examples (i.e., patients) described in terms of several descriptive variables and labelled with several target variables we want to learn a rule ensemble that will predict the values of target variables from the values of descriptive variables. In addition, we want to identify groups of patients that are similar in terms of all the target variables and desribe them in terms of descriptive variables.

Input data:

Given a subset of ADNI data with 916 patients described with 34 different variables, collected at baseline evaluation, we split the variables in two groups. In the first group we have descriptive variables: APOE4, Ventricles, Hippocampus, WholeBrain, Entorhinal, Fusiform, MidTemp, ICw, FDG, Aw45, CDRSB, ADAS13, MMSE, RAwLT_immediate, RAwLT_learning, RAwLT_forgetting, RAwLT_perc_forgetting, FAQ, MOCA. In the second group we have the target variables: EcogPtMem, EcogPtLang, EcogPtwisspat, EcogPtPlan, EcogPtOrgan, EcogPtDiwatt, EcogPtTotal, EcogSPMem, EcogSPLang, EcogSPwisspat, EcogSPPlan, EcogSPOrgan, EcogSPDiwatt, EcogSPTotal. We decide to learn a rule set with maximum two rules.

Example output:

y = $f(x)$ = (2.2, 1.8, 1.4, 1.5, 1.6, 1.9, 1.7, 2.2, 1.7, 1.5, 1.7, 1.8, 1.9, 1.8)

+ [**IF** (FAQ <= 5) *and* (ADAS13 <= 14) *and* (MidTemp > 16425) *# 512*

   THEN (-0.1, -0.1, -0.1, -0.1, -0.1, -0.1, -0.1, -0.2, -0.2, -0.1, -0.2, -0.2, -0.2, -0.2)]

+ [**IF** (FAQ <= 0) *and* (MOCA > 19) *and* (ADAS13 <= 15) *and* (APOE4 <= 1) *# 342*

   THEN (-0.1, -0.1, -0.1, -0.1, -0.1, -0.1, -0.1, -0.2, -0.1, -0.1, -0.2, -0.2, -0.2, -0.2)]

The resulting model is a sum of a constant vector (no linear equations were added to the model in this case) and two rules. Each of the rules can be interpreted as a cluster of examples that is described with associated conditions. The two clusters contain 512 and 342 patients, respectively.