Pot-luck challenge: TIED

Advanced Analytics, Intel, LTD
5000 W. Chandler Blvd
CH5-295
Chandler, AZ, 85226, USA

EUGENE.TUV@INTEL.COM

Editor: Isabelle Guyon, Dominik Janzing and Bernhard Schölkopf

Task(s) solved:

- Using training data, find all minimal sets of features with optimal predictivity
- For each of the feature set identified, build a classifier model of the target variable using training data and apply it to the testing data.

Method: Rule induction on relevant features

Feature selection method (ACE - Artificial Contrasts with Ensembles) was used to remove irrelevant features. Two rule induction techniques were used to find sets of features with optimal predictability: CART with surrogate splits and a supervised APRIORI. Both point to the same optimal sets of features.

- Feature selection: ACE is a combination of three ideas: A) Estimating variable importance using RF ensemble of trees of a fixed depth (3-6 levels) with the split weight re-estimation on OOB samples (gives more accurate and unbiased estimate of variable importance in each tree), B) comparing variable importance against artificially constructed noise variables using a formal statistical test, and C) Iteratively removing the effect of identified important variables to allow detection of less important variables. ACE method is outlined in (Tuv et al., 2006). The more comprehensive paper is submitted to JMLR (currently under review).

The results of ACE applied to the TIED dataset are shown on the Figure 1. The algorithm stopped after 3 iterations (no new relevant features found), and the resulting set of selected relevant (strongly and weakly) features is shown in the last column.

- Classification tree (Breiman et al., 1984) built on selected features shown on Figure1. Optimal tree has four terminal nodes, and gave CV BER $\sim 0.02$. The tree was used for the prediction on the test data. Figure 2 presents surrogate scores tables shown for each of the three splits. Note that for the first split on Column10 there are three surrogates with equivalent splits (Column1/2/3). Similarly for the second and the third splits equivalent splits are achieved by using Column11/12/13 and Column18/19/20 correspondingly.

- Supervised Apriori: we customized Apriori (Agrawal et al., 1993) algorithm to produce rules with known consequent - specific class of a categorical target. We use conditional support (fraction of the data from the specified class covered by the rule)

©2007 Author Name1 and Author Name2.
to dramatically simplify APRIORI rule tree construction. As a preprocessing step
numeric predictors are discretized, and levels of categorical predictors are optionally
clustered with respect to the target class using decision tree with MDL based pruning.
The preprocessing is done on each variable independently, and could result in suboptimal rules (this is the case for the target class=2, TIED). The set of the best rules
found by the algorithm is shown on Figures 3-4, and involve the same set of variables
\{1, 2, 3, 10\} \times \{11, 12, 13\} \times \{18, 19, 20\} found by a single tree (with surrogate splits).

**Implementation:** All the methods described above are implemented in C++ within Intel
Statistical Learning framework - IDEAL. It is not publicly available.

**Results:**

- Minimal sets of features with optimal predictivity: 36 sets of vars \(\rightarrow\) \{1, 2, 3, 10\} \times
  \{11, 12, 13\} \times \{18, 19, 20\}
- Model: Single 4-node classification tree built using any triple from the above cartesian
  product (see Figure 1) results in the equivalent model with CV BER \(\sim\) 0.02

**Keywords:** feature selection, tree classifier, rule induction, supervised Apriori

**References**

R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items
in large databases. In *Proceedings of the ACM SIGMOD International Conference on


E. Tuv, A. Borisov, and K. Torkkola. Feature selection using ensemble based ranking
against artificial contrasts. In *Proceedings of the International Joint Conference on Neural
Networks (IJCNN)*, 2006.
Figure 1: Left graph: The results of ACE applied to the TIED dataset. The algorithm stopped after 3 iterations (no new relevant features found), and the resulting set of selected relevant (strongly and weakly) features sorted by relative importance is shown in the last column. Right Graph: Classification tree built on the set of the relevant features identified by ACE. For each split surrogate scores are calculated for each variable (see the Figure 2).
Figure 2: Surrogate scores tables shown for each of three splits for the tree model built to classify TIED target. Note that for the first split on Column10 there are three surrogates with equivalent splits (Column1/2/3). Similarly for the second and the third splits equivalent splits are achieved by using Column11/12/13 and Column18/19/20 correspondingly.
Figure 3: Rules for the target class=0 (upper table). Perfect discrimination is achieved with one of the variables 1/2/3/10. Rules for the target class=3 (lower table). Perfect discrimination is achieved with one of the variables 11/12/13.
Figure 4: Rules for the target class=1 (upper table, a subset is shown). The best 36 equivalent rules found by the algorithm involve triples from the set \( \{1, 2, 3, 10\} \times \{11, 12, 13\} \times \{18, 19, 20\} \). Rules for the target class=2 (lower table). The best 9 equivalent rules found by the algorithm involve tuples from the set \( \{11, 12, 13\} \times \{18, 19, 20\} \).