Interpretable Classification

Besides accuracy an important aspect of every classifier is its interpretability. Especially in manufacturing we demand a transparent decision process. This provides insight into the data and is sometimes a formal requirement (e.g. for product quality control). Interpretability can be quantified through the sparsity level of the classifier, i.e., the number of relevant features. Sparse classifiers are guaranteed to yield good generalization performance.

Decision Lists

Decision lists are ordered collections of probabilistic rules. For binary classification \( (y = 0 \text{ vs } y = 1) \) their general form is:

\[
\begin{align*}
\text{IF } & \text{cond}_1 \text{ THEN } p_1 \\
\text{ELSEIF } & \text{cond}_2 \text{ THEN } p_2 \\
& \ldots \\
\text{ELSEIF } & \text{cond}_K \text{ THEN } p_K \\
\text{ELSE } & p_{K+1}
\end{align*}
\]

where \( p_k \in [0,1] \) are the probabilities for \( y = 1 \). The conditions are chosen from a pool of binary features of the kind attribute = value or attribute ∈ interval. More expressive power can be obtained by using conjunctions of multiple primitive features. Decision lists are similar to decision trees, but they use a linear representation rather than a hierarchy. Learning of decision lists is traditionally based on separate-and-conquer procedures (hill-climbing or beam search) which are extremely scalable algorithms that can be applied to very large datasets.

Directional Decision Lists

Greedy learning algorithms fail if they try to characterize the more complex class. By explicitly specifying the search direction we can relieve the algorithm from the responsibility to choose the direction of the rules. A directional decision list that characterizes the positive class \( (y = 1) \) is defined as satisfying

\[
P(y = 1 | \overline{\text{cond}_1}, \ldots, \overline{\text{cond}_{k-1}}, \text{cond}_k) \geq P(y = 1 | \overline{\text{cond}_1}, \ldots, \overline{\text{cond}_k})
\]

for all \( k = 1, \ldots, K \) (overline denotes negation). Directional decision lists are even more intuitive than general decision lists because the rules cannot alternate between explaining the two classes.

Monotone Probabilities

A strict subfamily of directional decision lists are decision lists with monotone probabilities, i.e., \( p_k \geq p_{k+1} \) for all \( k = 1, \ldots, K \). In many cases directional decision lists happen to have monotone probabilities. However, enforcing monotone probabilities with a greedy algorithm is difficult.

Classification

When used for classification, directional decision lists reduce to a sparse disjunction of conditions that approximate the desired class indicator

\[
1\{y = 1\} \approx 1\{\text{cond}_1 \lor \cdots \lor \text{cond}_K\}.
\]

This type of classifier is also known as a cascading classifier.

Simulated Data

On simulated data we empirically confirmed that the proposed family of directional decision lists is easier to train (in a greedy manner) than the unrestricted family of all decision lists.

Problem Symptoms in Manufacturing

We applied our method to industrial process measurements for manufactured fuel pumps. The dataset consisted of 100,000 samples with 2,000 features each (after discretization) and the task was to classify the parts as defective (NOK) or intact (OK). The selected discriminative features were mainly indicators for extreme values of the displaced volume under certain applied pressure levels.