Characterizing Model Errors and Differences

Stephen Bay and Michael Pazzani

ICML 2000

Presented by Charles Rosenberg

October 9, 2000

Paper Overview

Goal is to automatically characterize model errors or differences in the feature space of the problem. This characterization takes the form of a description:

> Classifier MC4 is 15% less accurate than average at predicting the salary of people whose occupation is craft-repair and are married. This represents 183 misclassified examples.

This is useful for:

1. identifying areas in the feature space where the model should not be trusted and may need improvement
2. providing insight into the problem space by identifying hard areas
3. understanding of changes to an algorithm affect its performance in the problem domain

STUCCO: A Contrast Set Miner

The STUCCO algorithm searches for contrast sets which are conjunctions of attribute-value pairs:

$$P(cset|G_1) \neq P(cset|G_2)$$

$$|support(cset, G_1) - support(cset, G_2)| \geq \delta$$

A χ^2 test is used to test for significant differences between populations.

An exhaustive search for all possible subsets is made starting with simpler hypotheses first.

In this application STUCCO searches for subpopulations whose accuracy differs significantly than the average.

Quantifying the Analysis

1. Effect Size: the number of additional examples that would have been correctly classified if the accuracy on the subpopulation was the same as the average

2. Rule Length: the number of terms in the conjunction used to identify an subpopulation

3. Stability: average agreement over training sets drawn from the same distribution:
   $agreement(X, Y) = \frac{|X \cap Y|}{|X \cup Y|}$

Results

Evaluation was performed on models induced from the Adult Census data set and the task was to predict which people had a salary of $50K / year or more.

	MC4 Errors		1NN vs. 5NN		NB vs. SP	
	C5	S	C5	S	C5	S
number of rules	52	143	685	123	46	192
median effect size	44	148	2	64	38	115
average rule size	2.9	2.0	3.8	2.1	2.6	2.6
stability	0.25	0.54	0.05	0.48	0.24	0.40

Table 4. Summary of C5 and Stucco (S) rule sets characterizing MC4 errors, comparing 1NN with 5NN and comparing Naive Bayes (NB) with SuperParent (SP).

Results

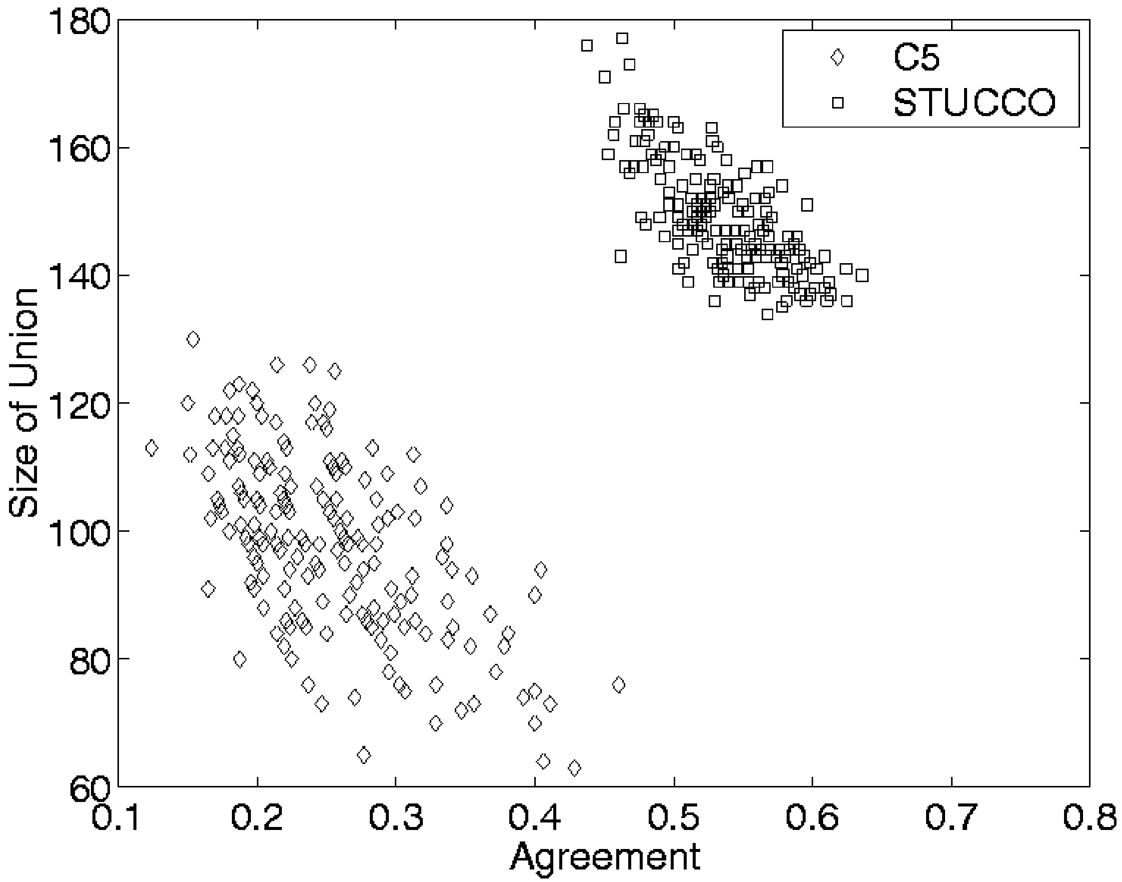

Figure 2. Stability of C5 and STUCCO