



Post-processing of Decision Trees



Alma Garcia
Almanza

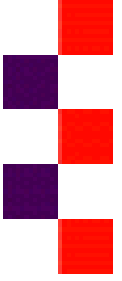


Edward Tsang

- Three methods developed for these objectives
- Based on machine learning and supervised learning
- Under the evolutionary paradigm
 - specifically Genetic Programming (GP)

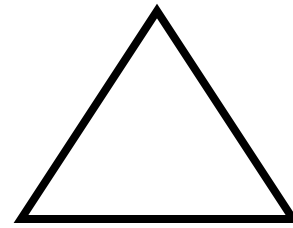
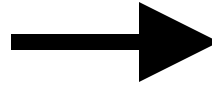
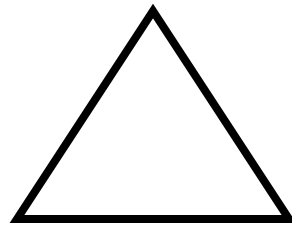
Objectives:

1. Improve precision and recall
2. Classify minority class in extreme imbalanced datasets
2. Produce a range of rules to suit user's preferences
3. Generate comprehensible solutions for user interaction.



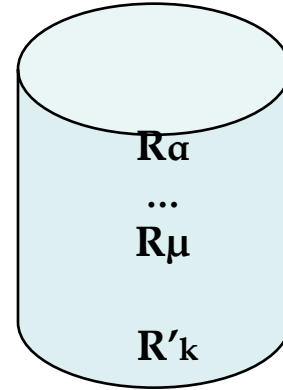
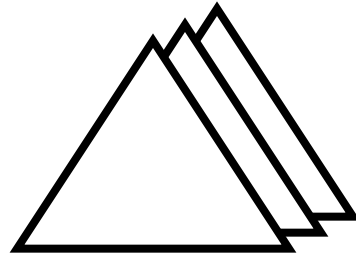
New Classification Methods

Scenario Method



(Pruning)

Repository Method



(Rules Collection)

Evolving Decision Rules

=

Repository Method

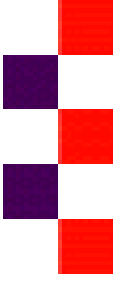
+

Evolution



Methods overview

<u>Repository Method</u> (RM)	<ol style="list-style-type: none">1. Classify minority class in imbalanced data sets2. Produce a range of classifications to suit the user's preferences	Our aim is to extract and collect different patterns that classify the positive cases (rare instances) in different ways.
<u>Evolving decision rules</u> (EDR)	<ol style="list-style-type: none">3. Provide understandable rules	EDR evolves a population of decision trees to form a repository of rules. The resulting rules are used to create a range of classifications
<u>Scenario Method</u> (SM)	<ol style="list-style-type: none">1. Analyze decision tree to detect and remove the rules that do not contribute to the classification task.	Scenario Method (SM) is a pruning procedure for decision trees created by GP. This pruning is based on the analysis of patterns in the decision tree.

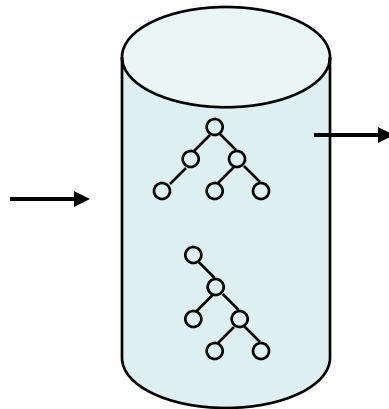


Repository Method

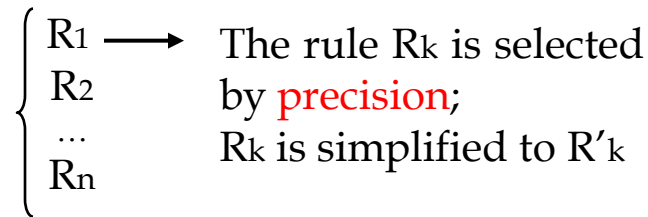
In order to mine the knowledge acquired by the evolutionary process Repository Method performs the following steps:

1- Rule extraction

Evolve a GP to create a population of decision trees

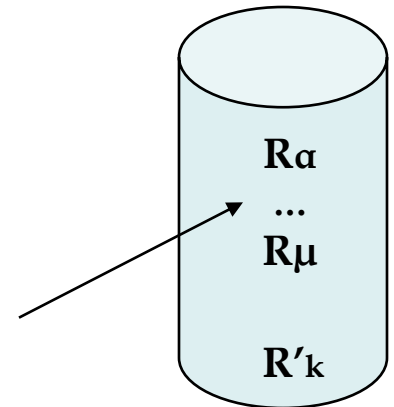


2- Rule simplification



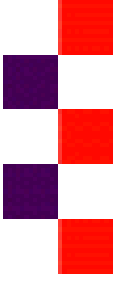
3- New rule detection

R'_k is compared to the rules in the repository by **similarity** (genotype)



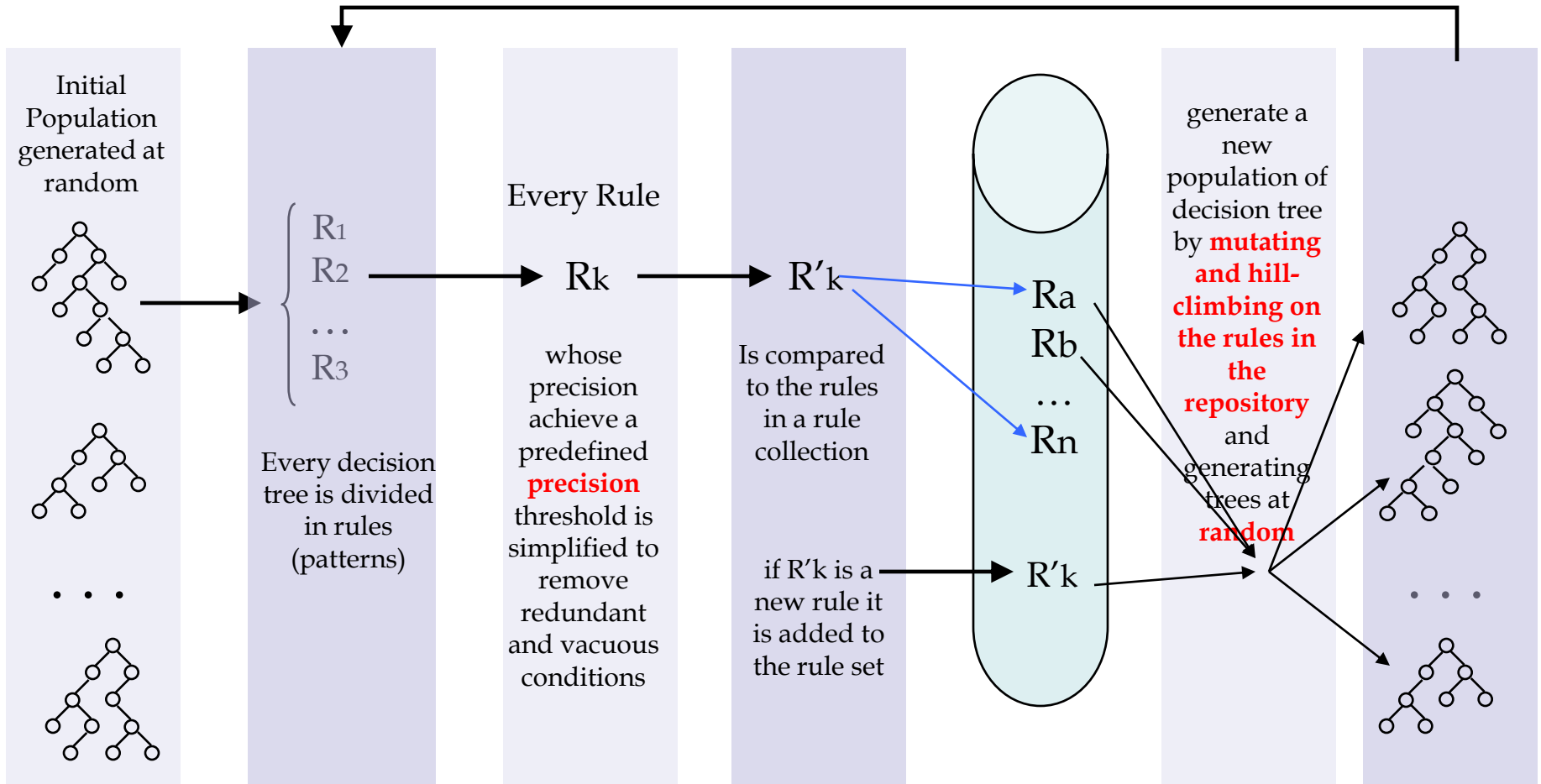
4- Add rule to the repository

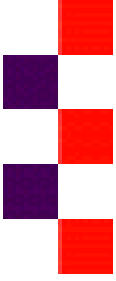
If R'_k is a new rule R'_k is added to the rule repository



Evolving Decision Trees

The new population is processed until the maximum number of iterations is reached



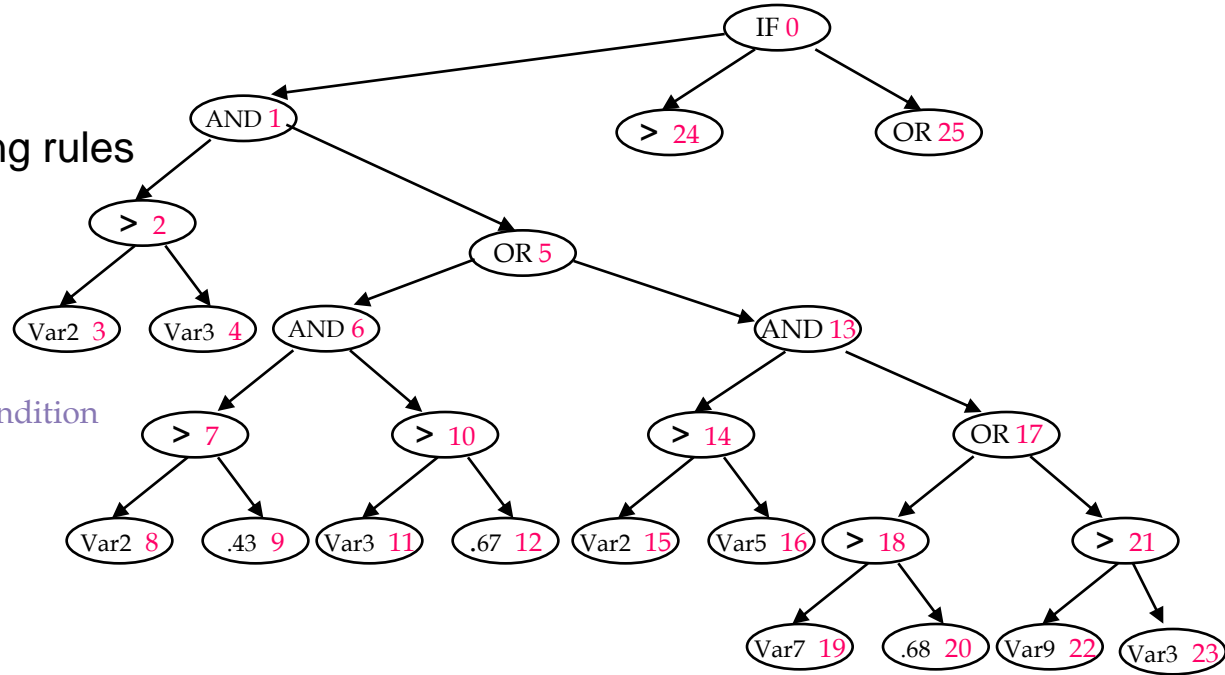


Scenario Method

The tree is composed by the following rules

R1 = { 2, 7, 10 }
R2 = { 2, 7, 10 }
R3 = { 2, 14, 21 }
R3 = { 2, 14, 21 }

Where the numbers represent the node of the condition



Procedure

- ❖ Evaluate every decision rule
- ❖ Consider that the R2 is not contributing to the classification task. Thus, we analyze every condition in R2 to determine which conditions are involved in other rules
- ❖ The only condition that is not involved in the other rules is 18
- ❖ Remove the condition