Actes IAF 2017

Admissible generalizations of examples as rules

Philippe Besnard[2] **Thomas Guyet**[3,1] **Véronique Masson**[4,1]

[1] LACODAM – Team Inria/IRISA-UMR6074, France
[2] CNRS/IRIT-UMR5505, France
[3] AGROCAMPUS-OUEST, France
[4] Rennes University-1, France

besnard@irit.fr thomas.guyet@irisa.fr veronique.masson@irisa.fr

Abstract

Rule learning is a data mining task consisting in generating predictive rules from a tabular dataset of examples labeled by a class identifier. We focus on propositional rule induction for which some representative algorithms are CN2 or Ripper. Such algorithms extract rules of the form "`IF` *Conditions* `THEN` *class-label*".

How to generalize examples is a central issue in a learning task. This work investigates the question of what is an admissible generalization of examples, to be called "rule elicitation", from an expert's viewpoint.

The overall rule learning process is viewed as a two-step process, depicted in Figure 1 :

1. Some subsets of data (some examples restricted to some attributes) are selected. In Figure 1, φ selects some possible subsets of data.
2. Each subset of data $(\mathcal{A}, S)$ is assumed to elicit (*i.e.*, is generalized by) a unique rule. The idea of the "eliciting" function f from $(\mathcal{A}, S)$ is to generate the rule π.

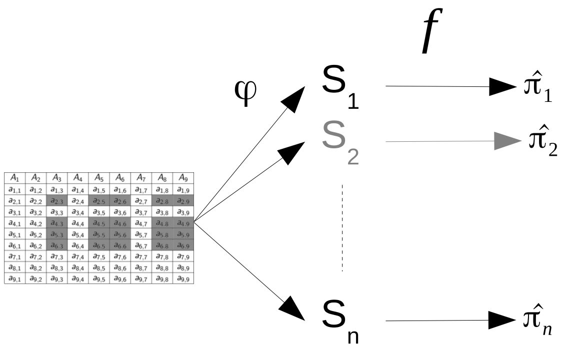

FIGURE 1 – Rule learning process: Abstract modeling

This article focuses on the f function, *i.e.*, how to elicit a rule from a subset of data. We do not tackle the question of determining φ, *i.e.*, how to select data subsets. It is assumed that rules are generated from all possible selected subsets. [1]

We introduce a formalization for generalization of examples as follows. A generalization of a set of examples S is an *admissible* rule $\pi = \widehat{S}$ where $\widehat{\cdot}$ is a closure-like operator. For any S and attribute A_i, generalizing $S_{\,i}$ (*i.e.*, S restricted to the A_i attribute) is identified with mapping $S_{\,i}$ to $\widehat{S}_{\,i}$ with properties of $\widehat{\cdot}$ taken from the list of Kuratowski axioms for closure or weaker versions of them.

We prove that some significant classes of generalizations are captured by preclosure and capping operators.

The main intuition is that rule admissible subsets of the range $\operatorname{Rng} A_i$ of an attribute A_i can be characterized as *choices* from the powerset of $\operatorname{Rng} A_i$. Depending on what principles underly the actual choice, a different kind of closure (*i.e.*, a weakening of Kuratowski's) embodies rule generation through generalization.

We explicitly give families of choice functions that induce preclosure, cumulation, and capping operators. The role of these functions is to choose an admissible rule among the set of supersets of the examples S.

We illustrate such concrete functions for numerical attributes and relate them to intuitive notions of value interpolation, respectively neighborhoods and intervals.

We check the behaviour of the CN2 algorithm against the notions introduced in our formalization. Our experiments seem to show that rules are generated only from the extreme values of example sets, regardless of the actual example distribution. This behaviour corresponds to operators stronger than capping (but still weaker than Kuratowski's). Yet, we should notice that these examples are specific cases of datasets with well-separated classes. In case of overlapping range of values, the rule choices that have been made depend on example distributions.

1. We have no practical objective, our purpose is not to design a new algorithm but to set a general framework to shed light on some characteristics of existing rule learning algorithms.