Fuzzy Decision Forest

Cezary Z. Janikow
and
Maciej Faifer

Department of Mathematics and Computer Science
University of Missouri – St. Louis
janikow@umsl.edu
faifer@arch.cs.umsl.edu

FID is a publicly available software for data classification. The methodology consists of three elements. First, there are two fuzzy partitioning methods, for partitioning attribute domains: one is top-down, creating minimal partitioning needed for generating a fuzzy tree, the other is bottom-up, creating more global partitioning. Second, there is a procedure for building fuzzy decision tree, which tree can also be interpreted as a set of fuzzy rules. The tree can be build using a number of potential fuzzy norms. Finally, there are a number of inference rules, for assigning classifications to new samples - based on the information extrapolated from the tree. The inferences fall into two basic categories: set-based (following local inferencing in fuzzy rules) and exemplar-based (following exemplar-based learning). The software can handle a mixture of features: symbolic, fuzzy terms, and numeric, and it can reason under incomplete/missing information. FID3.1 was presented at NAFIPS'99.

We are currently finalizing a major extension of the methodology - into fuzzy forests. This is intended to attack some traditional problems often associated with decision trees:

- decision trees are minimalistic in contained information, and they often degrade in complex domains with multi-dimensional relationships due to necessary "toss" situations when comparable alternatives are available
- this minimalistic approach fails when the necessary features (those used to build the tree) are missing

These problems have been addressed in the last few years in hybrid systems, in which a number of distinct trees were extracted and used with some voting rules. Fuzzy decision forest follows the same ideas, except that it uses alternatives, specific in local partitioning of the space. it extends the hybrid level into subdomains. In other words, while in those hybrid systems multiple choices are allowed at the level of the global search space, fuzzy decision forrest allows alternatives at every subspace level. Moreover, the choice of available alternatives is data and domain-driven. For example, if this is detected that the best attribute has a high probability of its features missing, alternatives will be provided even if they do not generate comparable quality partitioning.

Once built, fuzzy decision forest is used in classification inferences. If classification is performed using a one-dimensional fuzzy decision tree, evidence is combined and conflicts internal to leaves, and those across leaves, have to be resolved. If classification is performed using a forest, additional evidence is extracted from multiple dimensions in the forest, and then the additional conflict level is resolved. The software is in testing stages, but initial experiments indicate that fuzzy forests can greatly improve classification accuracy in more complex domains, and in particular when observations are noisy, uncertain, and/or missing. We will present some empirical comparisons during the conference.

FID is available at `http://www.cs.umsl.edu/~janikow/fid`.