The Iterated Version Space Algorithm
– A Machine Learning Approach

Jian Zhang, Ph.D. student
Dept. of Computer Science
University of Regina, Regina, Saskatchewan, S4S 0A2
jian@cs.uregina.ca

The Version Space Algorithm (VSA) [3] enables fast learning of a single concept if the training instances describe the concept consistently. That is, all of the input examples (positive or negative) have to be noise free. Unfortunately, in the real world, noise free training instances almost never exist. Most of the time, concept descriptions must be presented by a disjunctive set of hypotheses instead of a single hypothesis. Learning from inconsistent training instances is a common problem of concept learning. Many research efforts have been undertaken to produce algorithms with the ability to learn conjunctive and disjunctive hypotheses. This poster presents our recent research on the Iterated Version Space Algorithm (IVSA) [1, 2, 4, 5], which retains advantages of the Version Space Algorithm while handling disjunctive concepts.

Learning a disjunctive concept is similar to assembling a multi-dimensional jigsaw puzzle from a large selection of possible pieces. The target concept can be viewed as the puzzle, and an ordered list of disjuncts (called regional hypotheses [RH]) can be viewed as a set of pieces of the puzzle. One method to solve this problem is to repeatedly generate candidate pieces and try adding them to the puzzle until it is complete. With IVSA, the Hypothesis Generator produces candidate RHs (sets of puzzle pieces). Each RH covers some of input instances (part of the concept). The Hypothesis Assembler repeatedly picks the most promising hypotheses in each candidate RH according to a simple ranking heuristic and tries these hypotheses in each position in a list of accepted RHs. If adding a new hypothesis increases the coverage of the concept, it is placed in the position that causes the greatest increase; otherwise, this hypothesis is discarded. After all candidate RHs have been examined, the Hypothesis Remover takes every accepted hypothesis out of the list one at a time, to see if any of them can be removed without reducing accuracy. Then the process of hypothesis generation and assembly is repeated using, as input, the set of instances that are not yet covered. This learning cycle is repeated until further iterations cause no change in accuracy. The final list of accepted RHs is considered to be the concept description.

IVSA has been tested on a 20,000 word pronouncing dictionary. The experimental results show that IVSA is an efficient concept learning algorithm. It can produce disjunctive hypotheses and has good handling of inconsistent instances. The output concept description of the pronunciation matches the original pronunciation of the pronouncing dictionary with an accuracy of 95%.

References