

1023-68-321

Andrew G. Borden* (borden@wireweb.net), 8027 Robbins Glade, Converse, TX 78109.

Managing an NP-Complete Problem.

The NP-complete problem we are interested in solving is building a Bayesian classifier with certain properties. There are N classes of objects and there are M descriptors or parameters which distinguish between the classes. The descriptors may have either a continuous or discrete distribution. We have a data base providing statistical descriptions of each class in terms of their descriptors, or else we can build such a data base by aggregating discrete observations. There is a cost, perhaps time, for determining the value of each descriptor and we would like to minimize the average cost of making an identification or classification. This problem is NP-complete on the number of classes. We would like then, to build a classifier which:

- Provides classifications very efficiently, i.e. the mean time to classify is short
- Is self evaluating, i.e. provides a report on the probability of making a classification to a specified level of confidence, the average probability that a classification will be correct, and the average time required to make a classification

In order to solve this problem, we use Shannon's theory of information and Bayes Formula to build a demonstrably efficient, but sub-optimal decision tree. (Received September 05, 2006)