Genetic Algorithms For Classification and Feature Extraction

Min Pei, [1,2] Ying Ding, [2] William F. Punch, III, [3] and Erik D. Goodman [2]

[1] **Beijing Union University, Beijing, China**
[2] **Case Center for Computer-Aided Engineering and Manufacturing**
[3] **Intelligent Systems Laboratory, Department of Computer Science**

Michigan State University
Genetic Algorithms Research and Applications Group (GARAGe)
112 Engineering Building
East Lansing, MI 48824
Tel: (517)-353-4973. Fax: (517)-355-7516 e-mail: pei@egr.msu.edu

To be presented at the
1995 Annual Meeting,
Classification Society of North America,
June 22-25, 1995,
Denver, Colorado, USA

Genetic Algorithms For Classification and Feature Extraction

Abstract

In a decision-theoretic or statistical approach to pattern recognition, the classification or description of data is based on the set of data features used. Therefore, feature selection and extraction are crucial in optimizing performance, and strongly affect classifier design. Defining appropriate features requires interaction with experts in the application area. In practice, there is much noise and redundancy in most high dimensionality, complex patterns. Therefore, it is sometimes difficult even for experts to determine a minimum or optimum feature set. The "curse of dimensionality" becomes an annoying phenomenon in both statistical pattern recognition (SPR) and Artificial Neural Network (ANN) techniques. Researchers have discovered that many learning procedures lack the property of "scaling" i.e., these procedures either fail or produce unsatisfactory results when applied to problems of larger size.

To address this problem, we have developed two approaches, both based on genetic algorithms (GA's). The basic operation of these approaches utilizes a feedback linkage between feature evaluation and classification. That is, we carry out feature extraction (with dimensionality reduction) and classifier design simultaneously, through "genetic learning and evolution." The objective of these approaches is to find a reduced subset from the original *N* features such that useful class discriminatory information is included and redundant class information and/or noise is excluded. We take the following general approach. The data's original feature space is transformed into a new feature space with fewer features that (potentially) offer better separation of the pattern classes, which, in turn, improves the performance of the decision-making classifier. The criterion for optimality of the feature subset selected is usually the probability of misclassification. Since number of different subsets of *N* available features can be very large, exhaustive search is computationally infeasible and other methods must be examined. For example, a number of heuristic techniques have been used, but it is not clear under what circumstances any one heuristic should applied as each has its good and bad points. Genetic algorithms are good candidates for this task, since GAs are most useful in multiclass, high-dimensionality problems when heuristic knowledge is sparse or incomplete. In order to apply a GA to classification/discrimination tasks, we combined the GA with two different approaches: the K Nearest Neighbor decision rule and a production decision rule.

1) GA/KNN hybrid approach - Genetic algorithms combined with K Nearest Neighbor. Here the GA defines a population of weight vectors *W*, each of which is multiplied with every sample's data pattern vector *X*, yielding new feature vectors *Y* for the given data. The KNN algorithm then classifies the entire *Y*-transformed data set. The results of the classification are fed back as part of the fitness function, to guide the genetic learning or search for the best transformation. The GA "answer" then is the best vector *W* which yields the both fewest misclassifications and other characteristics, such as parsimony or cost of *W*. Thus the GA searches for an optimal transformation from pattern space to feature space for improving the performance of the KNN classifier.

2) GA/RULE approach - Genetic algorithms combined with a production rule system. This approach is based on the general structure of classifier systems, and focuses on the classification of binary feature patterns. A single, integrated rule format is used. The inductive learning procedure of the GA evolves a rule classifier using a known labeled "training" sample set. The result of training is a small set of "best" classification rules which classify the training set more accurately than other rule sets. By examining these "good" rules, one can determine the features which are most useful for classification. This provides information useful to domain experts for deriving new hypotheses.

The two methods described above were applied successfully to three high-dimensionality (96 or more features) biological binary pattern data sets with multiple classes. The results achieved indicate potential for application in many other fields. They can also be further developed using a GA combined with other classifiers, such as neural networks, to solve a fairly broad class of complex pattern recognition problems.