Incremental Hyperplane Partitioning for Classification

Tao Yang, Department of Computer Science and Software Engineering, Xi’an Jiaotong-Liverpool University, Suzhou, China

Sheng-Uei Guan, Department of Computer Science and Software Engineering, Xi’an Jiaotong-Liverpool University, Suzhou, China

Jinghao Song, Department of Computer Science and Software Engineering, Xi’an Jiaotong-Liverpool University, Suzhou, China

Binge Zheng, Department of Computer Science and Software Engineering, Xi’an Jiaotong-Liverpool University, Suzhou, China

Mengying Cao, Department of Computer Science and Software Engineering, Xi’an Jiaotong-Liverpool University, Suzhou, China

Tianlin Yu, Department of Computer Science and Software Engineering, Xi’an Jiaotong-Liverpool University, Suzhou, China

ABSTRACT

The authors propose an incremental hyperplane partitioning approach to classification. Hyperplanes that are close to the classification boundaries of a given problem are searched using an incremental approach based upon Genetic Algorithm (GA). A new method - Incremental Linear Encoding based Genetic Algorithm (ILEGA) is proposed to tackle the difficulty of classification problems caused by the complex pattern relationship and curse of dimensionality. The authors solve classification problems through a simple and flexible chromosome encoding scheme, where the partitioning rules are encoded by linear equations rather than If-Then rules. Moreover, an incremental approach combined with output portioning and pattern reduction is applied to cope with the curse of dimensionality. The algorithm is tested with six datasets. The experimental results show that ILEGA outperform in both lower- and higher-dimensional problems compared with the original GA.

Keywords: Classification, Genetic Algorithm, Hyperplane, Incremental Approach, Linear Encoding

1. INTRODUCTION

In machine learning field, Evolutionary Algorithms, Neural Networks, Fuzzy Logic and some other methods are widely used to derive solutions for classification problems which refer to assigning a given piece of input data into one of a given number of categories. Genetic Algorithms (GAs) are regarded as the typical algorithms among Evolutionary Algorithms. Moreover, among various kinds of Genetic Algorithm, rule-based GAs were utilized by
many researchers either in supervised or unsupervised learning (Lanzi et al., 2000). In the research work of Corcoran and Sen (1994), attributes in the static domain are classified with rule-based GA, which has achieved relatively low error rate. Thus, ILE-GA is designed with a linear encoding rule-based algorithm that solves problems with better flexibility. Furthermore, to reduce the complexity of deriving a reasonable solution for classification, an incremental approach is adopted for GA-based classifiers in a dynamic environment where training samples or new attributes may become available over time (Fangming Zhu & Guan, 2005). Similar incremental approach has also been exploited with artificial neural network to enhance the intelligence of the learnt models (Guan & Li, 2001). Incremental learning methods were proposed by Yamauchi et al. (1999) earlier for a different purpose, i.e. incremental pattern learning. In their approach, a small part of past learnt patterns will be learnt with new patterns. In contrast, the algorithm introduced in this paper will only incrementally add new experts and learnt patterns will not be learnt so that the training complexity will be reduced.

In this paper “Incremental Linear Encoding based Genetic Algorithm” (ILEGA) is designed to overcome the difficulties caused by either the complexity of pattern relationship or the rapid expansion of the solution space. All chromosomes are encoded by linear equations, which are more flexible compared with If-Then rules. This algorithm is effective in training complex solution structures and tackling the increase of problem dimension by utilizing an incremental approach.

We first elaborate the motivation of this algorithm in Section 2. Then the design and benefits of this algorithm are illustrated in Section 3. The experimental results on 2 artificial datasets and 4 benchmark datasets from UCI are reported in Section 4. The conclusions and suggestions for further studies are put forward in Section 5.

2. MOTIVATION

People from different background have been working on solving classification problems for years. However, many real-world classification problems, especially those with high-dimension, cannot yet be solved easily by conventional methods. Genetic algorithm (GA) is one of the intelligence algorithms proposed as a solution for classification problem during the past decades. Nevertheless, it still cannot provide high accuracy for some difficult classification problems. As that has been mentioned above, the difficulties of solving classification problems may come from either the complexity of the problem itself or the rapidly increased solution space. Therefore, in our analysis, the difficulty of a problem may derive from two aspects: first, people prefer to encode a set of If-Then rules into chromosomes when using GA to solve classification problems, which will lead to the formation of some rectangular subspaces in the pattern space. The problem will become harder when the complexity of spatial relationship between patterns increases using this encoding scheme. For instance, the problem will be easy to solve if the patterns can be partitioned by a set of hyper-planes that parallel to the axes, while it will become hard to solve if the patterns need to be partitioned by surfaces when using If-Then Rule encoded GA. On the other hand, the increase of problem dimensions will lead to the increase in problem complexity dramatically since with the increasing number of attributes, the solution space will also increase rapidly. This results in the significant growth of the problem difficulty. In the worst case, the initial fitness for all the chromosomes in the population will likely be zero due to the large solution space. As a consequence, the selector operator will have no clue in selecting better chromosomes and thus GA will degenerate to a random search algorithm. Due to all the reasons given above, ILEGa is proposed.
Related Content

Alignment: The Activity Domain in the Centre
[www.igi-global.com/chapter/alignment-activity-domain-centre/39681?camid=4v1a](www.igi-global.com/chapter/alignment-activity-domain-centre/39681?camid=4v1a)

Towards The New Episteme: Philosophy, Knowledge Science, Knowledge and Tacit Knowledge
[www.igi-global.com/article/towards-new-episteme/41732?camid=4v1a](www.igi-global.com/article/towards-new-episteme/41732?camid=4v1a)

Timely Autonomic Adaptation of Publish/Subscribe Middleware in Dynamic Environments
[www.igi-global.com/chapter/timely-autonomic-adaptation-publish-subscribe/68950?camid=4v1a](www.igi-global.com/chapter/timely-autonomic-adaptation-publish-subscribe/68950?camid=4v1a)
Multi-Objective Genetic Algorithm for Robust Clustering with Unknown Number of Clusters
www.igi-global.com/article/multi-objective-genetic-algorithm-robust/64211?camid=4v1a