Incremental Hyper-Sphere Partitioning for Classification

Jinghao Song, Department of Computer Science and Software Engineering, Liverpool University, Jiangsu, China

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

Binge Zheng, Department of Computer Science and Software Engineering, Liverpool University, Jiangsu, China

ABSTRACT

In this paper, an Incremental Hyper-Sphere Partitioning (IHSP) approach to classification on the basis of Incremental Linear Encoding Genetic Algorithm (ILEGA) is proposed. Hyper-spheres approximating boundaries to a given classification problem, are searched with an incremental approach based on a unique combination of genetic algorithm (GA), output partitioning and pattern reduction. ILEGA is used to cope with the difficulty of classification problems caused by the complex pattern relationship and curse of dimensionality. Classification problems are solved by a simple and flexible chromosome encoding scheme which is different from that was proposed in Incremental Hyper-plane Partitioning (IHPP) for classification. The algorithm is tested with 7 datasets. The experimental results show that IHSP performs better compared with those classified using hyper-planes and normal GA.

Keywords: Classification, Genetic Algorithm, Hypersphere, Incremental Hyper-Plane Partitioning (IHPP), Incremental Hyper-Sphere Partitioning (IHSP), Linear Encoding

1. INTRODUCTION

Evolutionary Algorithms, Neural Networks, Fuzzy Logic and many other methods were heavily used to generate solutions for classification and clustering problems. Genetic Algorithms are considered to be the most typical algorithms among Evolutionary Algorithms for classification problems which refer to assigning an input pattern into one of a given number of output categories. Furthermore, rule-based genetic algorithms were used by many researchers either in supervised or unsupervised learning (Lanzi et al., 2000). Corcoran and Sen (1994) classified the attributes in the static domain with rule-based GA, which has achieved a relatively low error rate. ILEGA (Yang, Guan & Song, 2013) solved classification problems with better flexibility using an incremental approach together with some innovative linear encoding rule. Moreover, such an incremental approach adopted for GA-based classifiers in a dynamic environment where training samples or new attributes may become available over time (Zhu & Guan, 2005), can obviously
reduce the complexity of deriving a reasonable solution for classification. A similar incremental approach has also been exploited with artificial neural network to enhance the learning performance (Guan & Li, 2001). Incremental learning methods were utilized 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 relevant with new patterns. In contrast, the algorithm introduced in this paper will only incrementally add new experts and learnt patterns will not be relearnt so that the training complexity will be reduced over time.

In the previous work, ILEGA with IHPP was proposed to solve classification problems with a higher classification rate compared with normal genetic algorithms. A classifier consisting of expert groups that are able to classify certain class of patterns, is served as the basis of the solution. Each expert group contains an unconstrained number of experts that are capable of dealing with a specific number of patterns in a specific class. Furthermore, a linear encoding scheme is applied in the specification of a chromosome so that the process of generating and decoding rules becomes simple, fast and flexible. Hyper-planes represented by linear equations mark the boundary of the patterns with more precision as the number of rules in every expert increases gradually. With the aid of genetic algorithm based optimization, several hyper-planes will form an expert which will be proficient in classifying a region enclosing certain class of patterns. The results of this approach has shown superiority to normal GA in many aspects.

ILEGA is again applied in this paper to overcome the difficulties which are caused by either the complexity of pattern relationship or rapid expansion of the solution space. Rather than encoding with If-Then rules in general expert systems, the chromosomes are still encoded with a linear encoding scheme, which is more flexible and robust. Together with IHSP, it can gain almost the same or better results in classification accuracy compared with IHPP and normal genetic algorithm. In addition, hyper-sphere partitioning generally has better performance in speed in most of the situations tested. Thus, IHSP is deemed to be effective and efficient.

We first elaborate the motivation of this algorithm in Section II. Then the design and benefits of this algorithm are illustrated in Section III. The experimental results on four artificial datasets and three benchmark datasets from UCI are reported in Section IV. The conclusions and suggestions for further studies are drawn in Section V.

2. MOTIVATION

Since the ILEGA with IHPP approach has solved many real world classification problems, the concept of ILEGA is exploited in this paper to make the most use of its advantages. However, for every expert in the training process of IHPP, up to 30 hyper-planes (rules) are encoded to classify the patterns. This has rendered the process of classification slow, because plenty of generations of chromosomes will be produced and validated. Although the generation of hyper-planes is fast, it is still unable to guarantee that all hyper-planes can form the boundary correctly in a short while, unless several more generations are derived. For problems with big data, this type of constraints can make the cost of training increased significantly. Nevertheless, for some problems with curved surfaces, e.g. patterns that form concentric circles, are not suitable for HPP to produce a satisfying solution.

As a matter of fact, hyper-sphere partitioning (HSP) approach has the capacity to overcome the difficulties mentioned above. For every expert, only one hyper-sphere is used to enclose patterns. So, every expert is distinct and no relationships between hyper-spheres need to be discovered. This simplifies the solution process and allows the solutions to be derived quickly. Additionally, the search space is reduced as only one centroid and the associated radius should be searched, which will rapidly accelerate the pace of producing hyper-spheres matching the criteria.
Related Content

**Theory of Cooperative Coevolution of Genes and Memes1**
[www.igi-global.com/chapter/theory-cooperative-coevolution-genes-memes1/28370?camid=4v1a](www.igi-global.com/chapter/theory-cooperative-coevolution-genes-memes1/28370?camid=4v1a)

**Improving the Efficiency of Color Image Segmentation using an Enhanced Clustering Methodology**

**Fractal Geometry as a Bridge between Realms**
[www.igi-global.com/chapter/fractal-geometry-bridge-between-realms/69455?camid=4v1a](www.igi-global.com/chapter/fractal-geometry-bridge-between-realms/69455?camid=4v1a)
A Heuristic Approach for Multi Objective Distribution Feeder Reconfiguration: Using Fuzzy Sets in Normalization of Objective Functions
www.igi-global.com/chapter/heuristic-approach-multi-objective-distribution/66818?camid=4v1a