Chapter 5
A Review on Evolutionary Prototype Selection: An Empirical Study of Performance and Efficiency

Salvador García
University of Granada, Spain

José Ramón Cano
University of Jaén, Spain

Francisco Herrera
University of Granada, Spain

ABSTRACT
Evolutionary algorithms have been successfully used in different data mining problems. Given that the prototype selection problem could be seen as a combinatorial problem, evolutionary algorithms have been used to solve it with promising results. This chapter presents an evolutionary data mining application known as evolutionary prototype selection. Various approaches have been proposed in the literature following two strategies on the use of evolutionary algorithms: general evolutionary models and models specific to prototype selection problem. In this chapter, the authors review the representative evolutionary prototype selection algorithms proposed, give their description and analyze their performance in terms of efficiency and effectiveness. They study their performance considering different sizes of the data sets, and analyze their behavior when the database scales up. The results are statistically contrasted in order to argue the benefits and drawbacks of each model.

INTRODUCTION
In supervised classification problems, we usually have a training set of samples (prototypes, instances) in which each example is labeled according to a given class. In the group of supervised classifiers, we can find the Nearest Neighbor (NN) rule method (Papadopoulos & Manolopoulos, 2004; Shakhnarovich...
et al., 2006) that predicts the class of a new prototype by computing a similarity measure (Pekalska et al., 2006) between this prototype and all prototypes from the training set, called the k-Nearest Neighbors (k-NN) classifier. In the k-NN classifier, k nearest neighbors vote to decide the class of the new instance to classify. A specific case is the 1-NN, where just the nearest neighbor is considered to take the decision.

Recent studies show that k-NN classifiers could be improved by employing numerous procedures. Among them, we could cite proposals on instance reduction (Liu & Motoda, 2002; Wilson & Martinez, 2000) for incorporating weights for improving classification (Paredes & Vidal, 2006), and for accelerating classification task (Gómez-Ballester et al., 2006), etc.

Prototype Selection (PS) is an instance reduction process consisting of maintaining those instances that are more relevant in the classification task of the k-NN algorithm and removing the redundant ones. This attempts to reduce the number of rows in the data set without any loss in classification accuracy and obtain an improvement in the classifier. Various approaches of PS algorithms have been proposed in the literature, see Wilson & Martinez (2000) and Grochowski & Jankowski (2004) for a review. Another process used for reducing the number of instances in the training data is the prototype generation, which consists of building new examples by combining or computing several metrics among the original data and including them into the subset of the training data (Lozano et al., 2006).

Evolutionary Algorithms (EAs) have been successfully used in different data mining problems (see Freitas, 2002; Ghosh & Jain, 2005; García-Pedrajas & Ortiz-Boyer, 2007). Given that PS problem could be regarded as a combinatorial problem, EAs have been used to solve it with promising results (Cano et al., 2003). We termed such approaches as Evolutionary Prototype Selection (EPS).

The increase in the numbers of training samples is a staple problem in PS (which is known as the Scaling Up Problem). This problem produces excessive storage requirement, increases time complexity, and affects generalization accuracy. These drawbacks are present in EPS too, because they result in an increment in chromosome size and execution time and also lead to a decrease in convergence capabilities of the E.A. Traditional EPS approaches generally suffer from excessively slow convergence between solutions because of their failure to exploit local information. This often limits the practicality of EAs on many large-scale problems where computational time is a crucial consideration. A first rapprochement on the use of EAs when this problem scales up can be found in Cano et al. (2005).

The aim of this paper is to present a review of the EPS algorithms proposed, provide their description and analyze their performance in terms of efficiency and effectiveness. We will study their performance considering different sizes of the data sets, aimed at studying their behavior when the database scales up. The results will be statistically contrasted in order to argue the benefits and drawbacks of each model.

This chapter is organized in the following manner. Section 2 presents the PS task formally and the scaling problem involved when the size of the data set increases. A review of EPS is given in Section 3. In Section 4 we describe the EPS methods studied. Details of empirical experiments and results obtained are reported in Section 5. Section 6 contains a brief summary of the work and the conclusions reached.

**BACKGROUND**

PS methods are instance selection methods (Liu & Motoda, 2002) which attempt to find training sets offering the best classification accuracy by using the nearest neighbor rule (1-NN), considering the distance measure as the Euclidean one.
Related Content

A Resource-Aware Dynamic Load-Balancing Parallelization Algorithm in a Farmer-Worker Environment
[www.igi-global.com/chapter/resource-aware-dynamic-load-balancing/68945?camid=4v1a](www.igi-global.com/chapter/resource-aware-dynamic-load-balancing/68945?camid=4v1a)

The Effects of a Knowledge Management-Based Model for Teacher Professional Development in Hands-On Approach
[www.igi-global.com/article/the-effects-of-a-knowledge-management-based-model-for-teacher-professional-development-in-hands-on-approach/125434?camid=4v1a](www.igi-global.com/article/the-effects-of-a-knowledge-management-based-model-for-teacher-professional-development-in-hands-on-approach/125434?camid=4v1a)

Strategic Management, Evolutionary Economics, and Complex Adaptive Systems
[www.igi-global.com/chapter/strategic-management-evolutionary-economics-complex/5139?camid=4v1a](www.igi-global.com/chapter/strategic-management-evolutionary-economics-complex/5139?camid=4v1a)

Principles of Classification
[www.igi-global.com/chapter/principles-classification/69672?camid=4v1a](www.igi-global.com/chapter/principles-classification/69672?camid=4v1a)