Evolutionary Mining for Image Classification Rules

Jerzy Korczak and Arnaud Quirin

LSIIT, CNRS/Université Louis Pasteur,
Pôle API, Boulevard Sébastien Brant F-67400 Illkirch cedex
{korczak,quirin}@lsiit.u-strasbg.fr

Abstract. In this article, an approach for creating image classification rules using evolutionary operators is described. Classification rules, discovered by application of a genetic algorithm on remote sensing data, are able to identify spectral classes with comparable accuracy to that of a human expert. Genetic operators and the fitness function are detailed, and then validated for hyperspectral images (more than 80 spectral bands). Particular attention is given to mutation operators and their efficiency in the creation of robust classification rules. In our case studies, the hyperspectral images contain voluminous, complex and frequently noisy data. The experiments have been carried out on remote sensing images covering zones of Lagoon of Venice and the city of Strasburg, France. It has been shown that the evolution-based process can not only detect and eliminate noisy spectral bands in remote sensing images but also produce comprehensive and simple rules which can be also applied to other images.

Keywords: Remote sensing image, classification rules, high resolution image, hyperspectral image, supervised learning, evolutionary learning, genetic algorithm.

1 Introduction

The design of robust and efficient image classification algorithms is one of the most important issues addressed by remote sensing image users. For many years, a great deal of effort has been devoted to generating new classification algorithms and to refine methods used to classify statistical data sets (Bock, Diday, 1999). At the time of this writing, relatively few workers in the machine learning community have considered how classification rules might be genetically discovered from raw and expertly classified images. In this paper, a new data-driven approach is proposed in order to discover classification rules using the paradigm of genetic evolution.

The unique source of information is a remote sensing image and its corresponding classification furnished by an expert. The images have been registered by various satellites (e.g. SPOT, LANDSAT, DIAS, ROSIS) that use different cameras having various spectral and spatial resolutions (Weber, 1995). These types of remote sensing images generally contain huge volumes of data, for instance an image of DAIS contains 79 bands of each one 2.8 Mbytes. And, sometimes they are very noisy due to coarse spatial resolution or unfavorable atmospheric conditions at the time the images are acquired. In addition, data may be also erroneous due to inexperienced operators of the measurement devices.

P. Liardet et al. (Eds.): EA 2003, LNCS 2936, pp. 153–165, 2004.

The aim of this research is to detail an evolutionary classification method applied to remote sensing images. More about evolutionary classifiers can be found in (DeJong, 1988) and (Ross, Gualtieri et al., 2002). As stated, the approach to discover classifiers is data-driven because the formulated classification rules are generated from data and are able to adapt themselves according to this available data, environment, and the evolution of classes. In remote sensing, the initial population of classification rules is randomly created from raw images and given classes, and then evolved by a genetic algorithm until the acceptable classification accuracy is reached.

In remote sensing literature, several classification approaches are presented, namely:

- pixel-by-pixel, each image pixel is analyzed independently of the others according to its spectral characteristic (Fjørtoft, Marthon et al., 1996),
- zone-by-zone, before classification, the pixels are aggregated into zones, the algorithms detect the borders of the zones, delimit them by their texture, or their repetitive patterns (Kurita, Otsu, 1993),
- by object, this is the highest level of recognition, the algorithms classify semantic objects, detect their forms, geometrical properties, spatio-temporal relations using domain knowledge (Korczak, Louis, 1999).

Our approach uses spectral reflectances; therefore, discovered classification rules are only able to find spectral classes rather than semantic ones. This spectral component of class description is essential to well recognize thematic classes. The approach has been validated using our software environment, called *I See You* (ICU). In this software, the object representation is not too sophisticated but it offers a high degree of freedom in description of symbolic expressions of rules and definition of genetic operators. The goal was to evaluate the capacity of the genetic approach to handle problems of over-generalization and over-fit in highly noisy and complex data. The ICU is a genetic-based classifier, where we have adapted and extended ideas of learning classifier systems, such as XCS (DeJong, 1988; Wilson, 1999), the s-classifiers, and „Fuzzy To Classify System“ (Rendon, 1997). We have also been inspired by the works of Riolo (Riolo, 1988) on gratification and penalization, and of Wilson (Wilson, 1999) on the exploration of the search space.

The paper is structured as follows. The basic concepts of image classification rules are introduced in Section 2. Section 3 details the discovery process of the classification rules. In this Section, the behavior of genetic algorithm functions is explained. Finally, two case studies on real remote sensing data are presented in Section 4.

2 Concept of Classification Rule Extracted from Remote Sensing Images

In general, classification rules are symbolic expressions and describe conditions to be held and actions to be taken if the conditions are satisfied. It must be underlined that in our approach the rules are discovered by an evolutionary process and are not given a priori by a domain expert.