Test Feature Classifiers and its Applications

V. Lashkia
Department of Inf. and Computer Eng.
Okayama University of Science
1-1 Ridai-cho, Okayama, 700-0005 Japan
lashkia@ice.ous.ac.jp

S. Aleshin
Department of Mathematical Theory
of Intellectual Systems
Moscow State University

Abstract

In this paper, we present a class of combinatorical-logical classifiers called test feature classifiers. Introducing kernels and a rejection option we discuss the properties and performance of the proposed classifiers. To test the performance of the classifiers, we apply them to a well-known phoneme database and textual region location problem. Our experimental results show that the proposed classifiers have very high stability and performance and suggest that they can be used in a variety of pattern recognition applications.

1. Introduction

Since the birth of electrical computing machines important advances have been made in the field of classifier learning from examples. However many classifiers in the learning phase require optimization methods, and have problems with convergence, stability and time efficiency. Statistical and structural methods learn badly when exact statistical or structural knowledge is not available. Any method based on metrics uses the hypothesis that the proximity in the data space generally expresses the membership of the same class, and therefore a data set which does not satisfy this condition cannot be treated by such an approach.

In this paper, we present a class of combinatorical-logical classifiers called test feature classifiers. These classifiers allow us to avoid many of the above drawbacks. Classifiers are generated directly from training samples using so-called *tests*, sets of features that are sufficient to distinguish patterns from different classes of training samples. The concept of the test was first introduced in [1] for the purpose of digital logic circuit analysis. It was then realized that tests could be very useful in pattern recognition. First, as the pattern recognition tool tests were used in [2]. The concept of test feature classifiers was presented in [3].

Test feature classifiers are m-degree polynomials, and can be used for partitioning the n-dimensional feature space, $m \leq n$, whose features are assumed to be binary-valued. Optimization methods, statistical, structural or metrical characteristics of patterns are not required. In this work, we have attempted to address some issues relevant to test feature classifiers which have not previously received attention. The performance of the proposed classifiers depend on the selection of tests and training samples. Introducing kernels and a rejection option we discuss the properties and performance of the proposed classifiers. We describe cases when a 100% of recognition rate can be achieved.

We applied the proposed classifiers to a well-known phoneme database and textual region location problem.

The choice of the phoneme database was dectated by two factors: the phoneme database is one of the largest available in the Internet, and it is known to be a difficult classification problem. We compere the performance of proposed classifiers to the performance of the conventional classifiers. Our simulations show that the proposed classifiers have much better performance than conventional methods.

Automatic textual region location deals with extracting image regions that just contain text. This is very useful in a number of applications such as image database retrieval, identification of products (such as books, CDs, etc), and reading street signs and notices which is an important part of the automatic function of robots and navigation systems. Several Approaches to text location in complex images have been proposed [7, 6, 8, 9]. In [7] a connected component approach in [8] a connected component approach in combination with OCR system have been proposed. In [6] a hybrid method has been proposed which combines the connected component approach and a spatial variance approach. In [9] the method of multivalued image decomposition was proposed. The connected compo-