Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition Letters 28 (2007) 136–143
Pattern Recognition Letters
www.elsevier.com/locate/patrec

Adaptive combination of adaptive classifiers for handwritten character recognition

Matti Aksela *, Jorma Laaksonen

Helsinki University of Technology, Adaptive Informatics Research Centre, P.O. Box 5400, FI-02015 HUT, Finland

Received 23 December 2005; received in revised form 14 April 2006
Available online 22 August 2006

Communicated by K. Tumer

Abstract

In this paper we examine the feasibility of combining two distinct layers of on-line adaptation for improving overall handwritten character recognition performance. These two approaches are adaptive classifiers and an adaptive committee used to combine them. On-line adaptive handwritten character classifiers are first discussed and the significant performance enhancements they can provide illustrated. We then examine the benefits from combining classifiers for this task, adaptive and non-adaptive, and present an adaptive committee structure suitable for this doubly adaptive framework. Experiments in combining the two adaptation approaches to form an adaptive committee consisting of adaptive member classifiers are described. The results show that while adaptation of the individual classifiers provides on average the most benefit in comparison to the non-adaptive reference level, the use of an adaptive combination of adaptive classifiers is still capable of enhancing the recognition performance by a significant margin. The usefulness of the proposed doubly adaptive approach is in this paper demonstrated in the domain of on-line handwritten character recognition, but we argue that the proposed methodology could also be applied to other application domains.

Keywords: Classifier combining; Adaptive classifiers; Adaptive committee; On-line adaptation; Handwritten character recognition

1. Introduction

In most pattern recognition tasks, the objective is to make as few errors as possible. There exist several approaches for improving the accuracy of a recognition system as a whole. In this paper, we will focus on two commonly used methods, namely on-line adaptation and classifier combining, which will be applied in the domain of on-line handwritten character recognition. These approaches are by no means mutually exclusive, on the contrary. They can effectively be used in combination with one another, as will be demonstrated here. Additionally, there is no reason for the methodology presented to be limited to only this particular application domain, as the adaptive committee structure is totally independent of the application and the classifier adaptation methods used are quite generic.

On-line adaptation to the specific classification task at hand is one performance improvement approach that can be very effective. This is especially true when a high level of intrinsic variation in the input data exists, but a substantial part of the variation can be explained by some underlying process or phenomenon. Examples of such tasks include speech and handwriting recognition, where the data from different users in general varies greatly, but each individual has a style that is reasonably consistent. It is this consistency that can be learnt during on-line use by starting with an user-independent system which is then adapted for optimal performance with that particular subject. With tasks exhibiting a high range of variability in the data, adaptation methods closely related to the data, for example

* Corresponding author. Tel.: +358 40 5872078; fax: +358 9 451 3277.
E-mail addresses: matti.aksela@hut.fi (M. Aksela), jorma.laaksonen@hut.fi (J. Laaksonen).

0167-8655/$ - see front matter
doi:10.1016/j.patrec.2006.06.016

methods which adapt the prototype set in a prototype-based classifier, are often more productive than more abstract or indirect techniques. Classifier adaptation has been implemented in such tasks for example by the use of a Gaussian model for cases with continuous style variation (Veeramachaneni and Nagy, 2003), by adaptive training of polynomial networks (Campbell and Broun, 2001), and by adaptation of the prototype set of a prototype-based classifier (Vuori et al., 2001).

Combining classifiers is another approach that has been shown to be useful on numerous occasions (Kittler et al., 1998). In general, the fundamental idea behind classifier combining is that the more different the mistakes made by the classifiers are, the more beneficial their combination can be Kuncheva et al. (2000) and Aksela and Laaksonen (2006), and, on the other hand, for most combination methods the level of obtainable benefit decreases as the similarity between the member classifiers increases. There exists a huge variety of classifier combination methods in use, for example voting schemes (Lam and Suen, 1994), Bayesian methods (Bouchaffra and Govindaraju, 1999), boosting (Drucker et al., 1993) and critic-driven combining (Miller and Yan, 1999), just to name a few.

Although most classifier combination rules are static by nature, also classifier combination strategies that are adaptive in some sense have been presented. Adaptation on the combiner level is commonly quite task-independent, as committees rarely deal with the input data directly. As such, they are easier to apply to a variety of cases when aiming for improved performance, even without detailed knowledge of either the task or the particular member classifiers. Examples of adaptive combination methods include the Adaptive Integration of Multiple Experts (AIME) system (Teow and Tan, 1995), an on-line learning boosting variant (Freund and Schapire, 1997), using a dynamic weighting coefficient predictor (Xiao et al., 2000), and the hierarchical Adaptive Combination of Classifiers (ACC) scheme (Mohan et al., 2001). One efficient adaptive committee strategy, the Class-Confidence Critic Combining (CCCC) committee, was introduced by the authors in (Aksela et al., 2003) for non-adaptive classifiers.

One may expect on-line adaptive approaches for both single and committee classifiers to provide significant benefits, as they are capable of learning during operation and becoming increasingly accurate with the task at hand. However, the downside to adaptive methods is the difficulty of operation under changing conditions precisely due to the adaptive methods' nature of learning from past behavior. It should still be noted that when constructing an adaptive system, special attention has to be paid to the over-learning issue – on-line adaptation commonly occurs at the expense of ability to generalize.

The main purpose of this paper is to examine if benefits from adaptive classifiers and adaptive committees could be combined to achieve even better performance through adaptive combination of adaptive member classifiers. This task is far from simple, as the adaptive nature of the classifiers themselves makes reliable adaptive combination much more difficult. In an adaptive system all the committee's decisions are based also on the member classifiers' previous performance. As the member classifiers' performances are constantly changing due to the classifiers' attempts to adapt to the data, predicting their behavior in the combination stage becomes very fragile. For this purpose a scheme of controlling the adaptivity through a weighted distance distribution model originally presented in (Aksela and Laaksonen, 2005) has been further developed and applied here.

We shall first describe a set of adaptive member classifiers for on-line handwritten character recognition in Section 2. Then in Section 3 we present an improved version of the adaptive committee introduced for static classifiers in (Aksela et al., 2003). Experiments are outlined in Section 4 and the results shown in Section 5. Through these experimental results we shall examine the feasibility of this doubly adaptive strategy and draw conclusions in Section 6.

2. The adaptive classifiers

The individual on-line handwritten character classifiers used in this study were introduced in (Vuori et al., 2001). The classifiers are based on calculating stroke-by-stroke distances between the input character and a set of prototypes. The number of coordinate points per character is not fixed, and hence a matching algorithm capable of dealing with curves consisting of varying numbers of data points must be used. The matching is here performed using dynamic time warping (DTW) (Sankoff and Kruskal, 1983). The classifiers then use the k-NN rule (Fix and Hodges, 1951) to find the most similar prototypes from the set of prototypes. As a by-product, the distances to the closest prototype in each class are collected.

DTW is used to compute one of three different distances, the point-to-point (PP), the normalized point-to-point (NPP) or the point-to-line (PL) distance. The PP distance uses the squared Euclidean distance between two data points as the cost function, whereas in the PL distance the points of a stroke are matched to lines interpolated between the successive points of the opposite stroke. The NPP distance is a normalized variation of the PP distance, where the sum of the matching costs is divided by the number of matchings made.

All character samples were scaled so that the length of the longer side of their bounding box was normalized and the aspect ratio kept unchanged. The centers of the characters were moved to the origin. For this we used two different approaches: the center of a character was defined either by its *mass center* or by its *bounding box center*. The classifiers themselves are described in more detail in (Vuori et al., 2001; Laaksonen et al., 1999) along with the methods used for the prototype set selection.

Two basic adaptation strategies namely *prototype addition* and *prototype modification*, have been used for these classifiers in a hybrid adaptation strategy found to be effective in (Vuori et al., 2001). In practice the adaptation of the