Available online at www.sciencedirect.com
SCIENCE DIRECT
Pattern Recognition 39 (2006) 1827–1838
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

Experimental study on prototype optimisation algorithms for prototype-based classification in vector spaces

M. Lozano[a,*], J.M. Sotoca[a], J.S. Sánchez[a], F. Pla[a], E. Pękalska[b,c], R.P.W. Duin[b]

[a] *Dept. Lenguajes y Sistemas Informáticos, Universitat Jaume I, Campus Riu Sec, 12071 Castellón, Spain*
[b] *Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands*
[c] *School of Computer Science, The University of Manchester, Manchester M13 9PL, UK*

Received 13 February 2006; accepted 6 April 2006

Abstract

Prototype-based classification relies on the distances between the examples to be classified and carefully chosen prototypes. A small set of prototypes is of interest to keep the computational complexity low, while maintaining high classification accuracy. An experimental study of some old and new prototype optimisation techniques is presented, in which the prototypes are either selected or generated from the given data. These condensing techniques are evaluated on real data, represented in vector spaces, by comparing their resulting reduction rates and classification performance.

Usually the determination of prototypes is studied in relation with the nearest neighbour rule. We will show that the use of more general dissimilarity-based classifiers can be more beneficial. An important point in our study is that the adaptive condensing schemes here discussed allow the user to choose the number of prototypes freely according to the needs. If such techniques are combined with linear dissimilarity-based classifiers, they provide the best trade-off of small condensed sets and high classification accuracy.

Keywords: Dissimilarity representation; Prototype selection; Adaptive condensing; EM algorithm; Normal density based classifier; Nearest neighbour rule

1. Introduction

An intuitive way of determining the class of an unknown object is by analysing its similarity to a set of prototypes either selected or generated from a given training set (TS) of objects with known class labels. In general, similarities or dissimilarities can be computed either from the raw object observations or based on an intermediate feature representation. A small set of prototypes has the advantage of a low computational cost and small storage requirements, while leading to similar, or even improved, classification performance. Various ways of designing a prototype set can be studied in Euclidean vector spaces. Two families of such optimisation procedures are editing and condensing.

Editing is the step in a learning process in charge of increasing the accuracy of predictions, when there is substantial noise in the training data. A basic editing algorithm removes noisy instances, as well as close border cases, eliminating a possible overlap between the regions from different classes and leaving smoother decision boundaries. Wilson introduced the first editing method [1]. Briefly, the k-nearest neighbour (k-NN) rule is used to estimate the class of each prototype in the TS and to remove those whose class labels do not agree with the ones judged by the k-NN rule. This algorithm tries to eliminate mislabelled objects from the TS as well as those near to the decision boundaries. Many researchers have addressed the problem of editing by proposing alternative schemes [2–5].

The *condensing* step aims at selecting a small subset of prototypes without a significant degradation in the classification accuracy. Two main groups of condensing techniques can be distinguished. These are the selective

* Corresponding author. Tel.: +34 9767 62109; fax: +34 9767 61914.
E-mail addresses: lozano@uji.es (M. Lozano), sotoca@uji.es (J.M. Sotoca), sanchez@uji.es (J.S. Sánchez), pla@uji.es (F. Pla), e.m.pekalska@tudelft.nl (E. Pękalska), r.p.w.duin@ieee.org (R.P.W. Duin).

0031-3203/$30.00
doi:10.1016/j.patcog.2006.04.005

schemes, which merely select a subset of the original training objects [2,6–9] and the adaptive schemes which modify them [10–15].

This paper discusses prototype optimisation methods, such as editing and condensing, for feature-based representations of classes in the context of prototype-based classification. Traditionally, the 1-NN rule is used for this purpose. It classifies objects based on the minimum distance to the given prototypes. Here, we show that the prototype sets, optimised to guarantee good performance of the 1-NN rule, lead to a higher classification accuracy, when more general dissimilarity-based classifiers are considered, as recently proposed [16,17]. These are weighted linear or quadratic combinations of the (Euclidean) distances computed between the test objects and the prototype sets found by dedicated condensing algorithms. Linear classifiers are especially of interest, since their computational complexity is comparable to that of the 1-NN rule.

Although the algorithms are run in vector spaces, the distances are by no means restricted to the Euclidean metric. Other distances, such as the l_p-distances, $d(\boldsymbol{x},\boldsymbol{y}) = (\sum_i |x_i - y_i|^p)^{1/p}$, $p > 0$ (metric for $p \geqslant 1$), or inner-product based distances, $d(\boldsymbol{x},\boldsymbol{y}) = 1 - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}/\|\boldsymbol{x}\|\|\boldsymbol{y}\|$, can be used. Moreover, the selective techniques can easily be applied to non-vectorial data (such as strings, graphs or shapes) provided that the pairwise dissimilarities are derived. This aspect has already been confirmed in earlier studies, in which numerous embeddings of metric and non-metric dissimilarity data (usually derived from non-vectorial representations) were analysed in Euclidean and pseudo-Euclidean spaces [16–18]. There, the prototypes were often chosen at random [18–21]. Other object selection techniques based on feature selection or clustering approaches were studied and appeared to be good for small prototype sets [16,17]. The conclusion of that work is that more general dissimilarity-based classifiers defined by the dissimilarities to a set of selected prototypes are competitive to the NN rule in the context of non-vectorial dissimilarity data.

In this paper we will confirm and extend this conclusion to vectorial representations, where (additionally to selection approaches) adaptive schemes can naturally be applied to *create* prototypes. These aspects have not been investigated so far. The reason behind our contribution is that the 1-NN (or k-NN) rule is often applied in vector spaces[1] for which editing-and-condensing techniques have been widely studied in order to reduce the computational burden while maintaining high accuracy. Consequently, our paper focusses on vectorial representations and the aspects related to the NN and condensed techniques. It is not our goal to present a general investigation into prototype selection techniques for dissimilarity data resulting from non-vectorial representations; in this case, the reader is referred to Refs. [16,17]. Instead, our goal is to study the applicability of condensing techniques, used in vector spaces to optimise the 1-NN rule, as prototype optimisation techniques for building more general dissimilarity-based classifiers.

We will show that linear or quadratic dissimilarity-based classifiers may successfully replace the 1-NN rule, especially when small sets of prototypes are needed. We will focus on Euclidean distances to maintain the connection with the traditionally used distances for selective condensing schemes in vector spaces. We will also compare these with adaptive reduction algorithms. Consequently, we focus on Euclidean distances in vector spaces and the use of condensing schemes determined in favour of the 1-NN rule. Adaptive schemes are especially of interest, as they can only be used when vectorial representations are available. Consequently, as they offer more flexibility, they should be more beneficial in vector spaces than the selective approaches.

Four condensing techniques are studied. Experiments are conducted to compare their ability to reduce the training size, while maintaining the discriminative power of the optimised prototypes. The classification performance is judged by the 1-NN rule and two more general dissimilarity-based classifiers.

The paper is organised as follows: Section 2 briefly reviews a number of condensing techniques which are used in our study. These are *MaxNCN* [14,23], *Reconsistent* [14], *LVQ* [24] and *MixtGauss* [15]. Section 3 briefly describes the framework of the prototype-based classification methods used for the evaluation of the derived condensed sets. In Section 4, the data sets are presented and the experiments are described, providing quantitative results and a further discussion. Finally, the main conclusions are summarised in Section 5.

2. Condensing methods to compare

Assume a TS of N instances, $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N\}$, representing J classes, $C = \{c_1, \ldots, c_J\}$. Each instance $\boldsymbol{x}_i$ is a point (a vector) in an n-dimensional feature space $\Re^n$. A condensed set consists of r, $r \ll n$, prototypes, which are either selected or generated from the examples of X. They are determined to represent efficiently the distributions of the classes and be well discriminative, when used to classify the training objects. Their cardinality should be sufficiently small to reduce both the storage and evaluation time.

A condensed set is said to be *consistent* with respect to a TS if the classification error, estimated by assigning all objects from the TS to the classes of their nearest neighbours in the condensed set, is small; see Refs. [3,7]. Therefore, given a set of prototypes representing the class distribution, the classification rate of the TS can be used to measure the consistency of this set. As the consistent condition is maintained, the smaller the number of prototypes in the condensed set, the better the final result is.

Four condensing algorithms are presented below. These are two selective schemes, *MaxNCN* and *Reconsistent*, and two adaptive schemes, *LVQ* and *MixtGauss*.

[1] This holds since the k-NN rule is known to be a simple and good classifier for large sample sizes thanks to its theoretical properties [22].