

Diversity measures for multiple classifier system analysis and design

Terry Windeatt *

Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, Surrey GU2 7XH, United Kingdom

Received 29 October 2003; received in revised form 4 March 2004; accepted 2 April 2004

Available online 18 May 2004

Abstract

In the context of Multiple Classifier Systems, diversity among base classifiers is known to be a necessary condition for improvement in ensemble performance. In this paper the ability of several pair-wise diversity measures to predict generalisation error is compared. A new pair-wise measure, which is computed between pairs of patterns rather than pairs of classifiers, is also proposed for two-class problems. It is shown experimentally that the proposed measure is well correlated with base classifier test error as base classifier complexity is systematically varied. However, correlation with unity-weighted sum and vote is shown to be weaker, demonstrating the difficulty in choosing base classifier complexity for optimal fusion. An alternative strategy based on weighted combination is also investigated and shown to be less sensitive to number of training epochs.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Decision level fusion; Multiple classifiers; Ensembles; Error-correcting; Binary coding

1. Introduction

A method of designing pattern recognition systems, known as the Multiple Classifier System (MCS) or committee/ensemble approach, has emerged over recent years to address the practical problem of designing classification systems with improved accuracy and efficiency. The aim is to design a composite system that outperforms any individual classifier by pooling together the decisions of all classifiers. The rationale is that it may be more difficult to optimise the design of a single complex classifier than to optimise the design of a combination of relatively simple classifiers.

Attempts to understand the effectiveness of the MCS framework have prompted the development of various measures. The Margin (Section 4.1) concept was used originally to help explain Boosting and Support Vector Machines. Bias and Variance (Section 4.2) are concepts from regression theory that have motivated modified definitions for 0/1 loss function for characterising Bagging and other ensemble techniques. Various diversity

measures (Section 3) have been studied with the intention of determining whether they correlate with ensemble accuracy. However, the question of whether the information available from any of these measures can be used to assist MCS design is open. Most commonly, MCS parameter values are set with the help of either a validation set or cross-validation techniques [1]. In [2] these measures are described and explained in the context of a vote counting framework. In this paper, in contrast to [2], the proposed measure relaxes the assumption on Hamming Distance (Eq. (22)) and is experimentally compared with various pair-wise diversity measures.

Although it is known that diversity among base classifiers is a necessary condition for improvement in ensemble performance, there is no general agreement about how to quantify the notion of diversity among a set of classifiers. Diversity measures can be categorised into two types [3], pair-wise and non-pair-wise. In order to apply pair-wise measures to finding overall diversity of a set of classifiers it is necessary to average over the set. Non-pair-wise measures attempt to measure diversity of a set of classifiers directly, based for example on Variance, entropy or proportion of classifiers that fail on randomly selected patterns. The main difficulty with

* Tel.: +44-1483-689286/259286; fax: +44-1483-686031.

E-mail address: t.windeatt@surrey.ac.uk (T. Windeatt).

diversity measures is the so-called accuracy–diversity dilemma. As explained in [4], as base classifiers approach the highest levels of accuracy, diversity must decrease so that it is expected that there will be a trade-off between diversity and accuracy. There has been no convincing theory or experimental study to suggest that there exists any measure that can reliably predict generalisation error of an ensemble. In [3] the desirability of using negatively correlated base classifiers in an ensemble is recognised, and it is shown experimentally that four pair-wise diversity measures (Eqs. (14)–(17)) are similarly related to majority vote accuracy when classifier dependency is systematically changed. The conclusion in [5] was that the Double Fault measure (Eq. (17)) showed reasonable correlation with some combination methods. Since there is a lack of a general theory on how diversity impacts ensemble performance, experimental studies provide an important contribution to discovering whether a relationship exists and if so whether it can be quantified and understood.

To be really useful for MCS design, a measure should be capable of extracting relevant information from the training set. Model selection from training data is known to require a built-in assumption, since realistic learning problems are in general ill-posed [6]. The assumption here is that base classifier complexity is varied over a suitable range and that over-fitting of the training set is detected by observing changes in diversity or correlation. It is shown experimentally in Section 6 that, over a range of datasets, some measures are well correlated with base classifier test error when number of training epochs is varied. As with Bias/Variance definitions (Section 4.2) one must assume that the underlying probability distributions are well-behaved, and it is easy to construct examples of probability distributions for which the method fails. The results in Section 6 also demonstrate that correlation with unity-weighting test error is not as strong as with the mean base classifier test error, illustrating the difficulty of choosing base classifier complexity for optimal fusion. An alternative strategy based on weighted combination is also investigated and the sensitivity of combined test error to number of epochs is compared with unity weighting.

The paper is organised as follows. A measure of correlation, based on a spectral representation of a Boolean function, is defined in Section 2. Conventional pair-wise measures are described in Section 3, which also includes proposal of a new pair-wise measure computed over pairs of patterns rather than pairs of classifiers. Margin and Bias/Variance are discussed in Section 4, and in Section 5 various weighted combination schemes are proposed. Experimental evidence, incorporating multi-layer perceptron (MLP) base classifiers in an MCS framework, is presented and evaluated in Section 6.

2. Spectral measure of correlation

The architecture envisaged in this paper is a simple MCS framework in which there are B parallel base classifiers whose outputs are combined either by voting or summation. Unity weighting will be designated as MAJ (majority vote) or SUM. The study is restricted to two-class supervised learning problems for which we assume that there are μ patterns with the label given to each pattern X_m denoted by $\omega_m = f(X_m)$ where $m = 1, \dots, \mu$ and f is the unknown function that maps X_m to the target label ω_m . In general the original features associated with each pattern are real-valued but we have no need to refer to them explicitly. Instead, we represent the m th pattern by the B -dimensional vector formed from the (real-valued) base classifier outputs ξ given by

$$X_m = (\xi_{m1}, \xi_{m2}, \dots, \xi_{mB}),$$

$$\xi_{mi} \in [0, 1], \quad \omega_m \in \{0, 1\}, \quad i = 1, \dots, B. \quad (1)$$

In this paper we stick to the convention that, where a pair of subscripts refers to pattern and classifier, the first subscript refers to the pattern and the second subscript to the classifier.

If one of two classes is assigned by each of B base classifiers, the m th pattern X_m in (1) may be represented as a vertex in the B -dimensional binary hypercube, resulting in a binary-to-binary mapping between classifier outputs and target labels

$$X_m = (x_{m1}, x_{m2}, \dots, x_{mB}),$$

$$x_{mi} \text{ and } \omega_m \in \{0, 1\}, \quad i = 1, \dots, B. \quad (2)$$

In [9], a spectral representation of a Boolean function $f(X)$ is proposed for characterising the mapping in (2) between base classifier outputs and target labels. In contrast to the conventional method of representing a Boolean function as single vertices of the binary hypercube, a spectral representation incorporates global information about the function in each spectral coefficient. A well known property of the transforms that characterise these mappings (e.g. Rademacher-Walsh transform [7]) is that the first order coefficients represent the correlation between $f(X)$ and x_i [8]. First order coefficients provide a unique identifier of $f(X)$ if it is linearly separable, and through table-lookup provide weights for a Threshold Logic Unit (TLU) implementation. The meaning of higher order coefficients as measures of correlation is given in [8].

In [2] a method of estimating the first order coefficients is described, which is based on representing $f(X)$ using a correlation measure called sensitivity σ that indicates whether a change in binary value x_i gives rise to a change in $f(X)$. For a completely specified Boolean function (truth table available), the m th pattern component x_{mj} is assigned σ_{mj} ($j = 1, 2, \dots, B$) as follows:

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات