Prototype Modelling of Body Surface Maps using Tangent Distance

Richard Patterson[1] and Gilbert MacKenzie

[1] Centre for Medical Statistics, Keele University, Keele, Staffordshire ST5 5BG, UK. E-mail: r.s.h.w.patterson@maths.keele.ac.uk

Abstract

In recent years the tangent distance approach of Simard et al (1993) to pattern recognition tasks such as handwritten character recognition has proven successful. Simard's distance measure could be made locally invariant to any set of transformations of the input and when implemented in one-nearest-neighbour classification of handwritten digits, outperformed all other classification schemes.
Hastie et al (1996) propose prototype models which generalise the concept of a centroid of a set of images in the Euclidian metric to a low-dimensional hyperplane in the tangent metric, and these prototypes can be used to reduce lookup time in classification.
We propose to apply and extend the tangent distance approach to classify a set of body surface maps, which are recordings of the electrical activity of the heart, of a large number of patients with various cardiac conditions. Using a grid of p electrodes attached to the anterior chest, we calculate a number of p-dimensional observation vectors for each patient and classify input maps on the basis of overall distance of map to prototypes derived from training set maps over all included observation vectors.

Keywords: Classification, Invariance Dimension, Prototype Models

1 Introduction

Many pattern recognition problems require allowance to be made for simple transformations of the input such as rotations, small scalings and location shifts. Simard et al (1993) addressed this problem in relation to handwritten character recognition. By generating a parameterised manifold for each image, a transformation-invariant metric could be defined as the minimum distance between two image manifolds. Approximating the manifold of an image by its tangent plane at the image itself led the distance measure being locally invariant to any set of transformations of the input and when

implemented in one-nearest-neighbour (1-NN) classification of handwritten digits, outperformed all other classification schemes.
Hastie et al (1996) proposed prototype models which generalise the concept of a centroid of a set of images in the Euclidian metric to a low-dimensional hyperplane defined by a point and a basis set of tangent vectors, chosen to minimise the average distance in the tangent metric from a subset of training images. These prototypes were used to greatly reduce lookup times in classification, with only a small increase in the rate of classification errors.
We propose to apply and extend the tangent distance and prototype approach to classify a set of body surface maps, which are recordings of the electrical activity of the heart, of a large number of cardiac patients. The 12 lead ECG device is used in hospitals throughout the world as the standard tool for diagnosing the condition of the heart. However, during the early stages of myocardial infarction, over one-fifth of patients will not have diagnostic ECG changes at the time of presentation, and these patients are therefore denied the potential benefits of early therapy and tend to have a higher mortality. Body surface mapping uses a larger array of electrodes to measure the spatial resolution of the heart's electrical activity on the chest wall. This extra information may facilitate the diagnosis in patients who are in the early stages of acute MI when diagnosis using the standard device is difficult. The set of maps to be classified includes control subjects with normal ECGs, patients with abnormal ECGs but no acute MI existing or developing, and patients with various sub-types of acute MI existing or developing at the time of mapping (including a proportion whose 12-lead ECG was unilluminatory at the time of presentation).

2 Tangent Distance

Consider the p-vector of observations arising from sampling an image at p standard locations. This pattern can be thought of as a point in p-dimensional space. When a pattern X is transformed in a single invariance dimension (e.g. rotated) which depends on a single parameter θ (e.g. the angle of the rotation), the set of all transformed patterns is a smooth one-dimensional curve $X(\theta)$ in p-dimensional space. In the case of multiple defined invariance dimensions with an n-dimensional parameter vector θ, $X(\theta)$ is a manifold of at most n dimensions. Unfortunately computation of these highly non-linear manifolds is impractical as generally they have no analytic expression.
The approach of Simard et al (1993) was to approximate small transformations of X by a tangent plane to $X(\theta)$ at point X, leading to the tangent model $\tilde{X}(\theta) = X + T(X)\theta$ where the columns of $T(X)$ are the tangent vectors for each invariance dimension, and tangent distance

$$D^T(X_i, X_j) = \min_{\theta_i, \theta_j} \parallel \tilde{X}_i(\theta_i) - \tilde{X}_j(\theta_j) \parallel$$

The approximation is valid locally, and therefore local invariance with respect to the set of transformations is induced.

3 Data Structure

Consider an electrode located on the chest at arbitrary co-ordinates (x, y). Let $f(t; x, y)$ be a time-dependent function which describes the electrical potential at co-ordinates (x, y) on the chest at time t. Then various iso-integral measurements of the form

$$z(t_a, t_b; x, y) = \int_{t_a}^{t_b} f(t; x, y)\, dt$$

and potential measurements of the form $f(t; x, y)$, where t is fixed, are calculated for each electrode. Using a grid of p electrodes attached to the anterior chest, we therefore calculate a number of p-dimensional observation vectors for each patient.

4 Prototype Models

The centroid $M = \sum_{i=1}^{N} X_i$ of a set of N points in p dimensions minimises the average squared norm i.e. M minimises $\sum_{i=1}^{N} \parallel X_i - M \parallel^2$. Hastie et al (1996) suggest some generalisations for this concept utilising the tangent metric. A point M can be found to minimise

$$\sum_{i=1}^{N} D^T(X_i, M)^2 = \sum_{i=1}^{N} \min_{\gamma_i, \theta_i} \parallel M + T(M)\gamma_i - X_i - T(X_i)\theta_i \parallel^2$$

where $T(M)$ is the basis for the tangent plane at the point M. Alternatively, the tangent subspace of point M can be included in the parameterisation. We then find M and V as the minimisers of

$$\sum_{i=1}^{N} \min_{\gamma_i, \theta_i} \parallel M + V\gamma_i - X_i - T(X_i)\theta_i \parallel^2$$

It is not necessary for V to have the same number of columns as $T(X)$; 0 to p columns are possible. One advantage of this approach is that the basis vectors for the tangent subspace of M may be viewed as class-specific invariances 'learned' from the data, while the effects of class non-specific invariances are removed. Other methods not involving tangent distance would mix the two sets of invariances.

Both schemes require solution by iterative algorithm, since the tangent basis vectors tend to be non-linear in M. In addition, it is necessary to incorporate the Singular Value Decomposition to find the hyperplane of given dimension closest in Euclidian distance to a set of points if we seek to calculate the tangent subspace of M. Hastie used such prototypes to drive a K-means algorithm and found they could greatly reduce classification lookup time with very little reduction in performance.

5 Results

The results from several classification schemes, utilising both Euclidean and tangent metrics, and nearest neighbour and prototyping approaches, were gathered and compared.
Here is a short summary of our results:
(a) Initially in the tangent metric paradigm, patterns were made invariant to rotations as well as scaling and location shifts in the x- and y- directions. This had a marked adverse effect on classification rates, however, and so the rotational invariance dimension was removed from the model.
(b) The tangent metric showed better classification performance than the Euclidean metric when used in conjunction with a 1-NN search; for the optimal number of nearest neighbours (seven) the results from the two methods were similar with a correct classification rate of 79% in both cases.
(c) Prototype models showed mixed results; the tangent centroid approach only gave 72% correct classification, while the tangent subspace method performed better with a rate of correct classification of 77%, only a decrease of 2% from the optimal result from nearest neighbour seaching with a very much reduced look-up time. In addition the subspace vectors generated have useful clinical interpretations (certain subspaces are linked to various sites of MI on the heart wall, etc.)
Prototyping via tangent subspace methods appears to be a useful method of reducing look-up times with the benefit of easier clinical interpretation but the penalty of slightly smaller correct classification rate. However in general the tangent metric ideas outlined do not appear to lead to any significant improvement in classification rates.
It may be the case that either our invariance dimensions or metric structure do not yet capture the nature of any systematic subtype variation.
We are examining the possibilities of including z (voltage) scale transforms and/or spatial covariance structures to augment our model.

References

Hastie, T., Simard, P.and Säckinger, E. (1996). Metrics and models for handwritten character recognition. In: *Proceedings of the 11th International Workshop on Statistical Modelling* .

Simard, P.Y., Le Cun, Y. and Denker, J. (1993). Efficient pattern recognition using a new transformation distance. In: *Advances in Neural Information Processing Systems* , Vol. 5, Morgan Kaufman, San Mateo, CA.