Supervised training technique for radial basis function neural networks

L. Bruzzone and D. Fernández Prieto

A novel supervised technique for training classifiers based on radial basis function (RBF) neural networks is presented. Unlike traditional techniques, this considers the class-membership of training samples to select the centres and widths of the kernel functions associated with the hidden units of an RBF network. Experiments carried out to solve an industrial visual inspection problem confirmed the effectiveness of the proposed technique.

Introduction: Classifiers based on radial basis function (RBF) neural networks [1] have been used in a wide variety of applications ranging from fault detection and diagnosis [2] to speech recognition [3]. The success of these classifiers stems from some of their properties (e.g. a fast training phase, systematic low responses for input patterns falling into regions of the input space where there are no training samples [1]) which render RBF neural classifiers effective tools in the area of pattern recognition for industrial applications. RBF neural classifiers consist of an input layer (with as many neurons as input features), a hidden layer, and an output layer (with as many neurons as classes to be recognised). The input neurons simply propagate the input features to the next layer. Each hidden neuron is associated with a symmetric kernel function (e.g. a Gaussian function) which provides a significant response only when the input falls into a localised region of the feature space. Each output neuron computes a simple weighted summation of the responses of the hidden neurons. The training of RBF networks is usually carried out in two steps: (i) the centres and widths of the kernel functions are selected (typically by using an unsupervised approach); (ii) the weights corresponding to the connections between the hidden units and the output units are fixed by means of a supervised algorithm (e.g. by minimising a sum-of-squares error function). A critical aspect of the training process is represented by efficient selection of the centres and widths of the kernel functions, as this choice may strongly affect errors made by the classifier [4]. In this Letter, we propose a simple, yet effective, supervised technique that selects these parameters by taking into account the class-membership information contained in the training set.

Proposed training technique: In RBF neural classifiers, the selection of the centres and widths of the kernel functions is generally carried out by applying a clustering technique (e.g. the k-mean-clustering algorithm [4]) to the whole training set, without considering the class-membership information about each training sample available in the training set. This may lead to the generation of mixed clusters containing data points belonging to different classes. The kernel functions generated by these clusters may reduce the separability of different classes in the kernel function space. This contributes to increasing the classification errors made by RBF neural classifiers. In addition, the random generation of mixed clusters when the number of hidden units is varied may result in an oscillatory behaviour of the classification error (i.e. small variations in the number of hidden units may cause large variations in the classification errors made by RBF neural classifiers). This problem makes the choice of the number of hidden units critical. The basic idea of the proposed technique is to carry out a clustering process on the training set by taking into account the class-memberships of the training samples in order to avoid the generation of mixed clusters. A detailed description of the technique is provided below.

Consider a pattern recognition problem in which each sample, described by an m-dimensional feature vector $x = (x_1, ..., x_m)$, is to be assigned to one of c different classes. Assume the availability of a training set $T = \{C_1 \cup C_2 \cup ... \cup C_c\}$, where C_i represents a subset of T containing all the training samples x^i_n, $(n = 1, ..., N_i)$ belonging to the class i. The proposed technique aims at partitioning the N_i samples of C_i into K_i disjoint subsets S^i_k so that at the end of the process $C_i = \{S^i_1 \cup S^i_2 \cup ... \cup S^i_{K_i}\}\ \forall\ C_i \subset T$. The algorithm involves the following steps:

Step 0: Let $i = 1$.

Step 1: For the class C_i the number of centres K_i is chosen according to both the number of training samples N_i and the dimension of the input space m [4].

Step 2: The k-means-clustering algorithm is applied to the subset C_i. The centres of the clusters $S^i_k \subset C_i (k = 1, ..., K_i)$ are initialised to different randomly chosen training samples belonging to C_i. Next each training sample belonging to C_i is assigned to the cluster nearest to it so that the sum of the Euclidean quadratic distances between the training points assigned to each cluster and the related cluster centre is minimised. This means that the algorithm finds a local minimum in

$$E^i_{K-means} = \sum_{k=1}^{K_i} \sum_{x^i_{k,j} \in S^i_k} \|x^i_{k,j} - \mu^i_k\|^2 \qquad (1)$$

where $x^i_{k,j}$= represents the jth training sample belonging to the cluster S^i_k, μ^i_k being the centre of the same cluster. When all the training samples included in C_i are assigned to different clusters so that $C_i = \{S^i_1 \cup S^i_2 \cup ... \cup S^i_{K_i}\}$, the centre of each cluster is updated by computing the new barycentre of the cluster. The procedure is iterated until convergence is reached [4].

Step 3: Each cluster centre μ^i_k is associated with one kernel function $\Phi^i_k(\mu^i_k, \sigma^i_k)$. Then, the p-nearest-neighbour [4] criterion is used to select the widths σ^i_k of the K_i kernel functions related to the class C_i. This technique makes use of the average distance from the centre of the kernel considered, $\Phi^i_k(\mu^i_k, \sigma^i_k)$, to the centres of the p nearest clusters included in C_i.

Step 4: Update $i = i + 1$. If $i \leq c$ go to *Step 1*; else *END* of the algorithm.

This algorithm allows us to obtain better separability of the different classes in the kernel function space. This reduces the classification errors made by RBF neural classifiers.

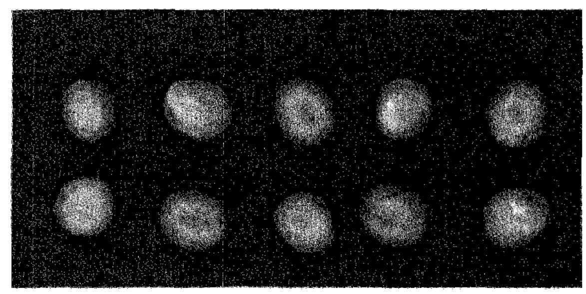

Fig. 1 *Example image of inspected peas*

Experimental results: The effectiveness of the proposed technique was assessed on an automatic visual-inspection system aimed at evaluating the quality of peas to be tinned at the end of an industrial production cycle. A camera, located on a conveyor that carried peas, acquired images in the visible and infrared spectral bands (Fig. 1). Eight features (related to colour, infrared radiation, shape, and dimension) were extracted from the images in order to characterise each sample. These features were given as an input to an RBF neural network classifier which was used to evaluate the quality of samples in terms of their appearance and spectral characteristics. This allowed the system to discriminate among the different qualities of peas for different commercial uses. Five classes, corresponding to different levels of product quality and to other objects that might be incidentally mixed with the peas on the conveyor, were defined: good, too ripe, and spotted peas, other

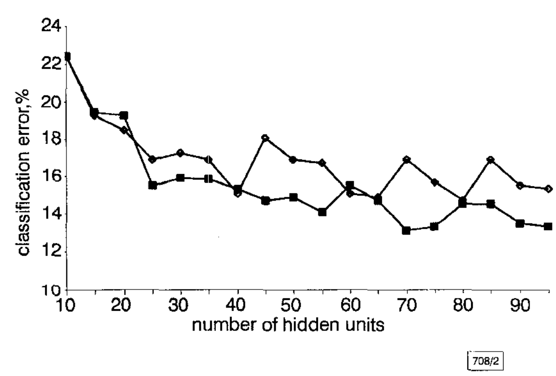

Fig. 2 *Classification errors on test set*

—◇— classical technique
—■— proposed technique

vegetable objects (e.g. small pieces of pod), and non-vegetable objects (e.g. small stones). In our experiments, we used two different RBF neural networks to associate each sample with one of the above classes: one was trained using the k-means clustering algorithm for the whole training set (i.e. without considering the class-memberships of the training samples); the other was trained using the proposed technique. The training phase was carried out on 505 samples; the test was performed on another 504 samples, independently chosen. Several trials were carried out by increasing the number of hidden neurons (and hence the number of kernel functions) from 10 to 100 (in steps of five). These trials allowed us to compare the behaviours of the classification errors made by the two training techniques for different numbers of kernel functions. For each trial carried out with the proposed technique, an equal number of kernel functions was chosen for each class.

Fig. 2 shows the classification errors made on the test set with the classical and the proposed techniques. The results confirm that the proposed technique significantly reduces the classification error made by the RBF neural classifier. In particular, the smallest classification error obtained using the classical technique was 14.70% (for 80 hidden units), whereas the minimum classification error obtained with the proposed technique was 13.06% (for 70 hidden units).

In addition, Fig. 2 confirms that the classification error incurred by the classical technique shows an oscillatory behaviour with regard to the number of hidden neurons considered (this makes it critical to fix the number of hidden units for an RBF neural classifier). On the contrary, the proposed technique results in a more stable trend of the classification error and consequently provides a better framework for choosing the architecture of an RBF neural classifier.

Conclusions: We have proposed a simple supervised technique for RBF neural network classifiers. In our experiments, this technique significantly reduced the classification error made by the classifier. In addition, a more stable behaviour of the classification error, with regard to the number of hidden units, was obtained. This renders the selection of the number of hidden units in an RBF network a less critical choice.

Acknowledgment: This research was partially supported by Bull HN Information Systems Italy S.p.A.

31 March 1998
Electronics Letters Online No: 19980789

L. Bruzzone and D. Fernández Prieto (*Department of Biophysical and Electronic Engineering, University of Genoa, Via Opera Pia 11a, I-16145 Genova, Italy*)

References

1 POWEL, M.J.D.: 'Radial basis functions for multivariate interpolation: a review' *in* MASON, J.C., and COX, M.G. (Eds.): 'Algorithms for approximation' (Clarendon Press, Oxford, 1987), pp. 143–167

2 PHILLIPS, W.J., TOSUNER, C., and ROBERTSON, W.: 'Speech recognition techniques using RBF networks'. Proc. IEEE WESCANEX 95. Communications, Power, and Computing, New York, USA, 1995, Vol. 1, pp. 185–190

3 NUNNARI, G., GALLO, A., REITANO, D., and OCCHIPINTI, L.: 'Neural nets in fault diagnosis', *Autom. Strum.*, 1997, **45**, (6), pp. 101–108

4 BISHOP, C.M.: 'Neural networks for pattern recognition' (Clarendon Press, Oxford, 1995)

10Gbit/s alternate polarisation soliton transmission over 300km step-index fibre link with no in-line control

P. Franco, A. Schiffini, R. Corsini, M. Romagnoli, M. Tamburrini and S. Cascelli

A straight-line in-field transmission was performed over a 300km step-index fibre link installed between Roma and Pomezia using a 10Gbit/s stream of 1550nm alternate-polarisation solitons. Standard SDH 10Gbit/s line terminals were used at the transmitter and receiver sides. Error-free transmission was obtained with optical amplifiers placed every 50km and with no in-line soliton control. It is important to underline that in this case the large chromatic dispersion was compensated for by fibre non-linearity without resorting to chromatic dispersion compensators.

The growing demand for large capacity communication links raises the problem of the already installed worldwide fibre infrastructure which is mostly constituted by step-index (SI) fibres with zero dispersion at 1.3μm. So far, dispersion management and/or wavelength division multiplexing (WDM) have been demonstrated to be suitable solutions for upgrading the existing fibre infrastructure. Those solutions require system adjustment, such as the introduction of dispersion compensating fibres or, in the case of WDM, an upgrade in terms of the number of channels, instead of simply increasing the capacity of the single channel, which would save in overall transmission bandwidth. In this Letter, we report an optimised transmission scheme which makes it possible to improve the maximum distance achievable on SI fibre with only a single channel, without resorting to the insertion of in-line components (such as chirped gratings, dispersion-compensating fibres, mid-span phase-conjugators, or even synchronous amplitude or phase modulators). The method simply exploits the reduced interaction efficiency experienced by orthogonal solitons to increase the duty cycle in the data stream [1], and it is implemented with a polarisation encoder which alternates orthogonal states of polarisation in the data stream. The effectiveness of alternate polarisation encoding has been theoretically predicted [2] and experimentally verified at 40Gbit/s over 800km in a dispersion-shifted fibre setup [3]. The maximum transmission distance achieved with our SI fibre system, based on alternate polarisation encoding, was 300km, thus extending the limit of 253km found with the dispersion supported transmission (DST) technique [4], another method of transmission with no in-line control.

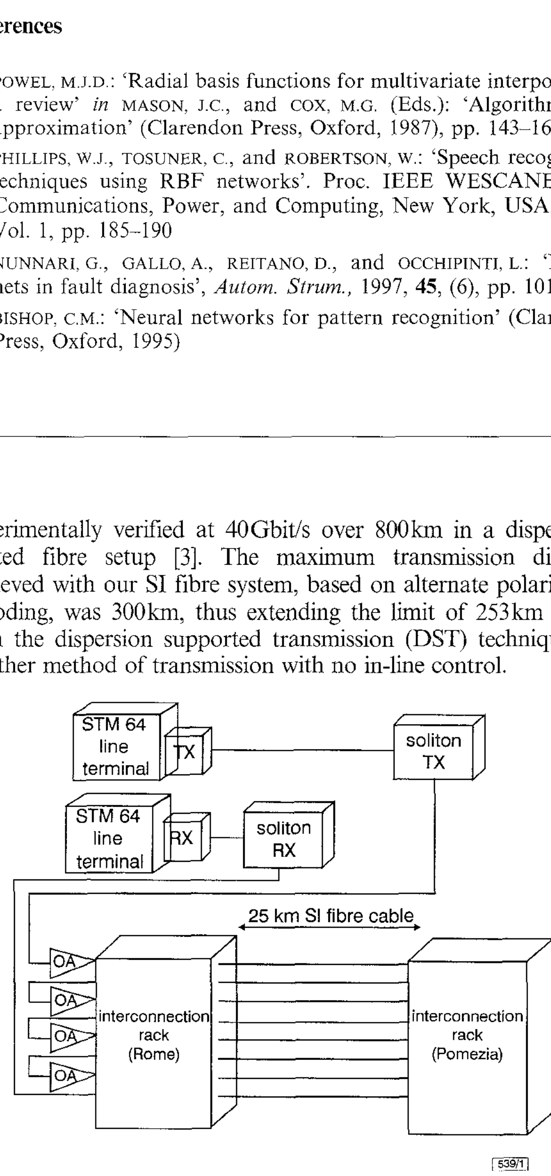

Fig. 1 *Line configuration*

The line configuration is shown in Fig. 1, with the experimental setup shown in Fig. 2*a*. The 5GHz soliton pulses emitted by the soliton generator have adjustable duration between 45 and 65ps and ~ –3dBm average power. The pulses are amplified by means of a polarisation maintaining optical amplifier up to 11dBm and sent to the polarisation encoder.

The 10Gbit/s electrical data are provided by the STM-64 line terminal whose clock also drives the entire setup. The electrical signal is divided into two 5Gbit/s electrical bit streams by means of the bit divider. The two 5Gbit/s electrical bit streams are used to modulate the 5GHz optical pulse streams by means of Mach-Zehnder modulators. Finally, the two 5Gbit/s optical pulse