Applied Soft Computing 23 (2014) 9–18

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

A radial basis function network classifier to maximise leave-one-out mutual information☆

Xia Hong[a], Sheng Chen[b,c,*], Abdulrohman Qatawneh[c], Khaled Daqrouq[c], Muntasir Sheikh[c], Ali Morfeq[c]

[a] School of Systems Engineering, University of Reading, Reading RG6 6AY, UK
[b] Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK
[c] Electrical & Computer Engineering Department, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia

ARTICLE INFO

Article history:
Received 27 June 2012
Received in revised form 29 January 2014
Accepted 4 June 2014
Available online 12 June 2014

Keywords:
Cross validation
Mutual information
Orthogonal forward selection
Radial basis function classifier

ABSTRACT

We develop an orthogonal forward selection (OFS) approach to construct radial basis function (RBF) network classifiers for two-class problems. Our approach integrates several concepts in probabilistic modelling, including cross validation, mutual information and Bayesian hyperparameter fitting. At each stage of the OFS procedure, one model term is selected by maximising the leave-one-out mutual information (LOOMI) between the classifier's predicted class labels and the true class labels. We derive the formula of LOOMI within the OFS framework so that the LOOMI can be evaluated efficiently for model term selection. Furthermore, a Bayesian procedure of hyperparameter fitting is also integrated into the each stage of the OFS to infer the l^2-norm based local regularisation parameter from the data. Since each forward stage is effectively fitting of a one-variable model, this task is very fast. The classifier construction procedure is automatically terminated without the need of using additional stopping criterion to yield very sparse RBF classifiers with excellent classification generalisation performance, which is particular useful for the noisy data sets with highly overlapping class distribution. A number of benchmark examples are employed to demonstrate the effectiveness of our proposed approach.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Model evaluation in terms of good generalisation performance is essential in the development and analysis of data-based learning algorithms for the construction of object classifiers. A fundamental concept in the evaluation of model generalisation capability is that of cross validation [1]. For example, in regression application, leave-one-out (LOO) cross validation is often used to estimate generalisation error by choosing amongst different model architectures [1]. In general, cross validation is required in most algorithms for model generalisation evaluation, and this often contributes significantly to computational cost for many model paradigms. Luckily for the linear-in-the-parameters models, the LOO cross validation can be exercised without actually splitting the training data set and estimating the associated models, by making use of the Sherman–Morrison–Woodbury theorem [2].

Moreover, for the linear-in-the-parameters models, the orthogonal least squares (OLS) based forward selection algorithm can efficiently construct parsimonious models [3,4], and has been a popular learning tool for associative neural networks, such as radial basis function (RBF) networks [5], fuzzy and neuro-fuzzy systems [6,7] as well as wavelets neural networks [8,9]. The OLS algorithm for RBF network learning [5] has also been utilised in a wide range of engineering applications, including aircraft gas turbine modelling [10], fuzzy control of multi-input multi-output nonlinear systems [11], power system control [12], fault detection [13], electric arc furnace load modelling [14], macromodelling of nonlinear digital I/O drivers [15], real-time power dispatch [16], fine tracking of NASA's 70-m-deep space network antennas [17], identification of urinary tract infection [18], stent reendothelialization [19], taxonomy and remote sensing of leaf mass per area [20], and many more.

For regression applications, regularisation methods based on a penalty function on l^2-norms of the model parameters are developed to carry out parameter estimation and model structure selection simultaneously [21–27]. From the powerful Bayesian

☆ X. Hong acknowledges the support of the UK EPSRC. This paper was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, under grant no. (1-4-1432/HiCi). The authors, therefore, acknowledge with thanks the DSR technical and financial support.

* Corresponding author.
E-mail addresses: x.hong@reading.ac.uk (X. Hong), sqc@ecs.soton.ac.uk (S. Chen), qatawneh@kau.edu.sa (A. Qatawneh), haleddaq@yahoo.com (K. Daqrouq), mshaikh@kau.edu.sa (M. Sheikh), morfeq@kau.edu.sa (A. Morfeq).

http://dx.doi.org/10.1016/j.asoc.2014.06.003
1568-4946/© 2014 Elsevier B.V. All rights reserved.

learning viewpoint, it can be shown that for linear-in-the-parameters models this parameter regularisation is equivalent to a maximised *a posteriori* probability (MAP) estimate of the parameters by adopting a Gaussian prior for the model parameters [22,24–28]. Furthermore, a regularisation parameter is equivalent to the ratio of the related hyperparameter to the noise parameter, lending to an iterative evidence procedure for solving the optimal regularisation parameters [24–28]. Note that, with the OLS algorithm, the evidence procedure for updating regularisation parameters becomes particularly efficient [22,25–27].

In information theory, the mutual information (MI) between two random variables is a quantity that measures the mutual dependence of the two variables [29,30]. The MI measure, as a fundamental measure in communications, has also been extensively used in regression applications, such as nonlinear system modelling [31,32], and pattern recognition applications, such as the feature selection [33], the registration of medical images [34] and gene classifications [35]. Note that in the existing literature MI criteria are normally used for training regression models or classifiers. Naturally if the MI is used as model structure selection metrics for classifier design, there is still the need to address model generalisation issue.

Against this background, in this work we propose to construct two-class RBF classifiers using the orthogonal forward selection (OFS) scheme, which selects one model term at each stage of the construction procedure by maximising the leave-one-out mutual information (LOOMI) between the classifier's predicted class labels and the true class labels, as well as incorporates a Bayesian procedure of hyperparameter fitting to efficiently derive the regularisation parameters. The paper contains two elements of novel contribution. Firstly, an original derivation of analytically evaluating the LOOMI efficiently is introduced, which facilitates the automatic model structure selection process with no need of using a predetermined error tolerance to terminate the forward selection process. Secondly, a novel Bayesian framework of calculating local regularisation parameters is designed specifically for the forward selection process, which leads to a very sparse classifier. Classification results for a number of benchmark examples demonstrate that our proposed approach efficiently construct very sparse RBF classifiers with excellent generalisation performance.

It is worthy emphasising that our contributions are significant. In the existing literature, the MI is used for training regression models and classifiers, but not used for model structure selection by optimising model generalisation capability. Instead of focusing on the usual training performance, to the best of our knowledge, our work is the first one that applies the MI for the effective model structure determination by introducing the novel LOOMI to incrementally maximise the classifier's model generalisation capability directly. Bayesian regularisation is also a well-known and widely used technique, e.g. in the support vector machine (SVM) and the relevance vector machine (RVM) [24] as well as in our previous orthogonal forward selection (OFS) based learning algorithms [22,25–27]. All these existing Bayesian regularisation approaches however involve an iterative procedure for updating the set of regularisation parameters. Specifically, given the values of all the regularisation parameters, model selection is carried out, and the resulting model is then used to update the set of regularisation parameters. This procedure iterates until both the selected model and the set of regularisation parameters converge. In this study, we introduce a novel Bayesian analysis for local regularisation parameter selection effectively nested within the OFS step. More particularly, each OFS stage also effectively fits one regularisation parameter from the data and this task is computationally very fast. Thus there is no need for iteratively performing the model selection and fitting the regularisation parameters several times. This paper is organised as follows. Section 2 introduces the two-class classifier construction using the OFS procedure and the concept of mutual information. In Section 3, we introduce model selection based on fast computing of the LOOMI. In Section 4, we carry out a Bayesian analysis for local regularisation parameter selection nested within the forward selection step. Section 5 presents the complete OFS algorithm that integrates joint parameter estimation with Bayesian regularisation and LOOMI model term selection. In Section 6, experimental results are employed to demonstrate the effectiveness of our proposed approach. Our conclusions are given in Section 7.

2. RBF classifier and mutual information

Consider the N labelled training data samples that belong to an approximately balanced two-class data set, denoted as $D_N = \{\boldsymbol{x}(k), y(k)\}_{k=1}^N$, where $\boldsymbol{x}(k) = [x_1(k)x_2(k)\cdots x_m(k)]^{\mathrm{T}} \in \mathbb{R}^m$ are m-dimensional feature vectors, and $y(k) \in \{\pm 1\}$ is the class type of $\boldsymbol{x}(k)$. We use the data set D_N to construct a RBF classifier of the form

$$\begin{cases} \tilde{y}^{(M)}(k) = \mathrm{sgn}(\hat{y}^{(M)}(k)), \\ \hat{y}^{(M)}(k) = f^{(M)}(\boldsymbol{x}(k)) = \displaystyle\sum_{i=1}^{M} \theta_i \phi_i(\boldsymbol{x}(k)), \end{cases} \tag{1}$$

where

$$\mathrm{sgn}(y) = \begin{cases} -1, & y \le 0, \\ 1, & y > 0, \end{cases} \tag{2}$$

$\tilde{y}^{(M)}(k)$ is the estimated class label for $\boldsymbol{x}(k)$ based on the M-term RBF model output $\hat{y}^{(M)}(k)$, and M is total number of regressors or model terms, while θ_i are the model weights, and the regressor $\phi_i(\boldsymbol{x})$ takes the form of Gaussian basis function given by

$$\phi_i(\boldsymbol{x}) = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{c}_i\|^2}{\tau}\right) \tag{3}$$

in which $\boldsymbol{c}_i = [c_{1,i}c_{2,i}\cdots c_{m,i}]^{\mathrm{T}}$ is the centre vector of the ith RBF unit and $\tau > 0$ is a RBF width parameter. We assume that each RBF unit is placed on a training data, namely, all the RBF centre vectors $\boldsymbol{c}_i$ are selected from the training data $\{\boldsymbol{x}(k)\}_{k=1}^N$, and the RBF width τ has been predetermined, for example, using cross validation.

Denote $e^{(M)}(k) = y(k) - \hat{y}^{(M)}(k)$ as the M-term modelling error for the data point $\boldsymbol{x}(k)$. Over the training data set D_N, further denote $\boldsymbol{y} = [y(1)y(2)\cdots y(N)]^{\mathrm{T}}$, $\boldsymbol{e}^{(M)} = [e^{(M)}(1)e^{(M)}(2)\cdots e^{(M)}(N)]^{\mathrm{T}}$, and $\boldsymbol{\Phi}_M = [\boldsymbol{\phi}_1\boldsymbol{\phi}_2\cdots\boldsymbol{\phi}_M]$ with $\boldsymbol{\phi}_l = [\phi_l(\boldsymbol{x}(1))\phi_l(\boldsymbol{x}(2))\cdots\phi_l(\boldsymbol{x}(N))]^{\mathrm{T}}$, $1 \le l \le M$. We have the M-term model in the matrix form of

$$\boldsymbol{y} = \boldsymbol{\Phi}_M\boldsymbol{\theta}_M + \mathbf{e}^{(M)}. \tag{4}$$

Here $\boldsymbol{\theta}_M = [\theta_1\theta_2\cdots\theta_M]^{\mathrm{T}}$. Let an orthogonal decomposition of the regression matrix $\boldsymbol{\Phi}_M$ be

$$\boldsymbol{\Phi}_M = \boldsymbol{W}_M\boldsymbol{A}_M, \tag{5}$$

where

$$\boldsymbol{A}_M = \begin{bmatrix} 1 & a_{1,2} & \cdots & a_{1,M} \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{M-1,M} \\ 0 & \cdots & 0 & 1 \end{bmatrix} \tag{6}$$

and

$$\boldsymbol{W}_M = [\boldsymbol{w}_1\boldsymbol{w}_2\cdots\boldsymbol{w}_M] \tag{7}$$

Download English Version:

https://daneshyari.com/en/article/495389

Download Persian Version:

https://daneshyari.com/article/495389

Daneshyari.com