Neurocomputing 182 (2016) 145–153

Contents lists available at ScienceDirect

Neurocomputing

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

Brief Papers

Dempster–Shafer theory-based robust least squares support vector machine for stochastic modelling

Chuang Zhou, XinJiang Lu *, MingHui Huang

State Key Laboratory of High Performance Complex Manufacturing, School of Mechanical & Electrical Engineering, Central South University, Hunan 410083, China

A R T I C L E I N F O

Article history:
Received 18 June 2015
Received in revised form
28 November 2015
Accepted 28 November 2015
Communicated by Wei Chiang Hong
Available online 17 December 2015

Keywords:
Modelling
LS-SVM
D–S theory
Robustness
Noise
Fuzzy clustering

A B S T R A C T

Noise can be produced from various types of sources with different spectral distributions. This often causes the least squares support vector machine (LS-SVM) to be less effective since the LS-SVM is sensitive to noisy data. In this work, a Dempster–Shafer (D–S) theory-based robust LS-SVM is proposed, which has a more reliable modelling performance under various noise regimes. A distributed LS-SVM is first developed to construct the evidence data set. Fuzzy clustering is then used to construct an evidence base from the data. D–S theory is further used to fuse different pieces of evidence to derive the parameters for the construction of a robust LS-SVM. This robust model can represent the original system well even in the presence of different types of random noise. Case studies are presented to demonstrate the effectiveness of the proposed LS-SVM approach.

1. Introduction

The least squares support vector machine (LS-SVM) is computationally attractive as it is only required to find the solutions of a set of linear equations [1–4]. It has become a popular data-driven modelling method and has been applied successfully across a range of end-uses [1–6].

Usually, there is noise within the data and it often comes from a variety of sources with different distributions [7–9]. Sampling errors, modelling errors, measurement errors, and operation errors all serve to complicate matters. This often causes problems in the robust modelling of the LS-SVM [10–12] because the least squares method is sensitive to outliers and is non-optimal in the case of non-Gaussian noise or noise distributions with heavy tails.

Most discussions about the LS-SVM are based on the assumption that the training data are not corrupted by noise [6,13]. In contrast, there are also many studies into ways to reduce the effect of noise on the LS-SVM model [12,14,15]. A range of loss functions have been employed to improve robustness such as, the weighted SVM [10], the reweighting SVM [16,17] and Nesterov's SVM (NESVM) [18]. In addition, some authors have contributed to the design of probabilistic SVM methods, which incorporate probabilistic information about the noise in their modelling. For example, the confidence intervals of the LS-SVM for regression have been derived [19]. Bayesian inference methods have also been applied to take into consideration probabilistic information in SVM classification and modelling [1,20]. Recently, a probabilistic SVM was proposed to enhance its functioning in noisy environments [6,13]. However, these probabilistic SVM methods must have a priori knowledge of the probability distribution of the noise or estimate it from the available data. This is often difficult to obtain when a complex system undergoes various kinds of transition, which alters the nature of the random noise. In addition, most of them only work well for noise that has a Gaussian distribution [15]. Due to these difficulties, a robust LS-SVM is still needed in order to permit a more reliable modelling performance in various kinds of random noise conditions.

Dempster–Shafer (D–S) theory is useful for handling uncertainty and imprecision [21–26]. With respect to more classical approaches to inference, such as the classical Bayesian approach, the use of D–S theory avoids the necessity of assigning prior probabilities (which would be extremely difficult to estimate) and provides more intuitive tools for managing uncertain knowledge [27]. D–S theory is also meant to avoid classical conditional probability limitations in the combination of items of evidence [28]. The most important benefit of D–S theory is the absence of an assumption about the independence of focal elements representing evidence [29]. Although many studies have combined D–S theory and SVMs for the fusion of multiple information sources and their classification [30–32], they have rarely been applied for

* Corresponding author.
E-mail address: luxj@csu.edu.cn (X. Lu).

http://dx.doi.org/10.1016/j.neucom.2015.11.081
0925-2312/

robust modelling LS-SVM in noisy environments. This is particularly the case when that noise comes from various sources with different distributions.

In this work, a D–S theory based robust LS-SVM (DS-RLSSVM) is proposed to model an unknown system with various kinds of random noise. The paper is organised as follows: the statement of the problem is given in Section 2; Section 3 presents details of the robust LS-SVM; Section 4 provides verification of the simulation, while conclusions are drawn in Section 5.

2. Problem description

An unknown system with various kinds of random noise may be represented as

$$y=f(x,\varepsilon_1,\cdots,\varepsilon_n) \tag{1}$$

where x and y are the input and output of the system respectively, f is an unknown function, and $\varepsilon_1,\ldots,\varepsilon_n$ are various types of random noise with different distributions.

This unknown system may be modelled by the following LS-SVM from its samples:

$$\tilde{y}=w^T\varphi(x)+b \tag{2}$$

here w and b are model parameter vectors and φ is a mapping to a high-dimensional and potentially infinite-dimensional feature space.

Given a training set $\{x_i,y_i\}_{i=1}^n$, the following optimisation problem in normal LS-SVM modelling is formulated [1]:

$$\min_{w,e_i,b} J(w,b,e)=\min_{w,e_i,b^2}\tfrac{1}{2}\|w\|^2+\tfrac{\gamma}{2}\sum_{i=1}^{n}e_i{}^2$$

$$s.t.\ y_i=w^T\varphi(x_i)+b+e_i,\ i=1,\ldots,n \tag{3}$$

where e is the modelling error and γ is the regularisation parameter. To solve this optimisation problem, a Lagrangian function is developed

$$\Gamma(w,b,e;a)=J(w,b,e)-\sum_{i=1}^{n}a_i\{w^T\varphi(x_i)+b+e_i-y_i\} \tag{4}$$

where a_i $(i=1,\ldots,n)$ are Lagrange multipliers. The conditions for optimality are given by

$$\frac{\partial\Gamma}{\partial w}=0,\quad \frac{\partial\Gamma}{\partial b}=0,\quad \frac{\partial\Gamma}{\partial e_i}=0,\quad \frac{\partial\Gamma}{\partial a_i}=0 \tag{5}$$

By solving Eq. (5), parameters a_i and b are obtained as follows:

$$\begin{bmatrix} 0 & 1_n^T \\ 1_n^T & \Omega+\frac{1}{\gamma}I_n \end{bmatrix}\begin{bmatrix} b \\ a \end{bmatrix}=\begin{bmatrix} 0 \\ Y \end{bmatrix} \tag{6}$$

where $Y=[y_1\cdots y_n]^T$, $1_n=[1\cdots 1]^T$, $a=[a_1\cdots a_n]^T$, $\Omega_{ij}=K(x_j,x_i)=\varphi(x_j)^T\varphi(x_i)$ with a positive definite kernel $K(\cdot,\cdot)$.

The resulting LS-SVM model for function estimation becomes

$$\tilde{y}(x)=\sum_{i=1}^{n}a_iK(x,x_i)+b \tag{7}$$

with $b=\dfrac{1_n^T\left(\Omega+\frac{1}{\gamma}I_n\right)^{-1}Y}{1_n^T\left(\Omega+\frac{1}{\gamma}I_n\right)^{-1}1_n}$, $a=\left(\Omega+\dfrac{1}{\gamma}I_n\right)^{-1}(Y-1_nb)$

This conventional approach to LS-SVM modelling is very sensitive to noise as shown in Fig. 1. Since the parameters a and b are identified from samples, the stochastic nature of the original system is incorporated into these parameters. For example, under each uncertainty $\eta_i=[\varepsilon_{1i},\ldots,\varepsilon_{ni}]$, there will be a corresponding LS-SVM model with the corresponding parameters a_i and b_i. All uncertainties will correspond to a group of parameters $\{(a_1,b_1),(a_2,b_2),\ldots,(a_m,b_m)\}$. Although the normal LS-SVM is optimal if the training samples are affected by Gaussian noise, it is less effective for random noise that has a different distribution [15]. Hence, the choice of robust parameters (a,b) is a key issue for LS-SVM under these conditions.

3. D–S theory-based robust LS-SVM

A robust LS-SVM modelling method is proposed in Fig. 2, which is developed for use in unknown, nonlinear systems with various kinds of random noise. This method first mines the data for information about the real system that is free from noise. To derive the robust LS-SVM parameters, it further uses evidence inference to fuse together different elements of evidence information, upon which a robust LS-SVM model is constructed. Since this approach may guarantee the model built here robust to noise, it may effectively model an unknown system with various kinds of random noise.

The detailed configuration of this robust LS-SVM approach, also called a D–S theory-based robust LS-SVM (DS-RLSSVM), is shown in Fig. 3. This method combines the advantages of distributed LS-SVM, fuzzy clustering and D–S theory. Integration of the distributed LS-SVM and fuzzy clustering is done to construct evidence information, while D–S theory is used to fuse evidence information to derive robust LS-SVM parameters. This method incorporates the following key functions:

- Evidence estimation: a distributed LS-SVM is proposed to mine evidence from the system data, upon which fuzzy clustering is used subsequently to construct the evidence.
- Evidence inference: D–S theory is used to fuse different pieces of evidence to estimate the parameters of the robust LS-SVM. From that, the robust LS-SVM model is then derived. This model should be robust to noise and represents the original system accurately.

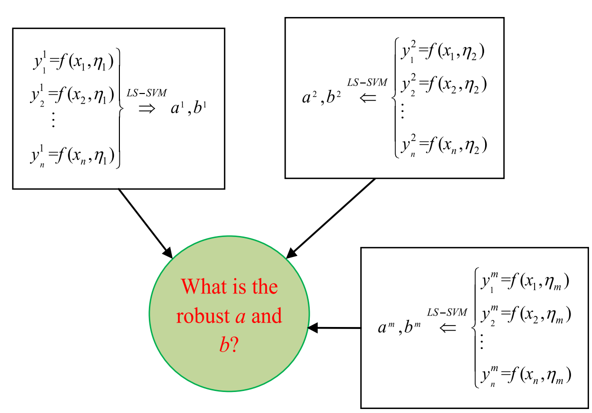

Fig. 1. Effect of noise on the LS-SVM parameters.

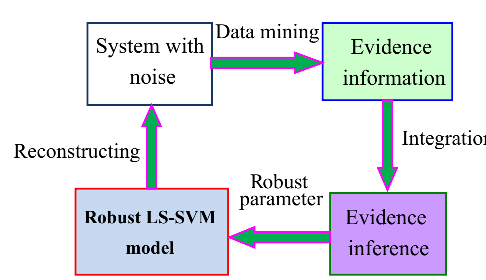

Fig. 2. Robust LS-SVM modelling method.