Available online at www.sciencedirect.com
SCIENCE DIRECT®
Fuzzy Sets and Systems 157 (2006) 1092–1113
FUZZY sets and systems
www.elsevier.com/locate/fss

On support vector regression machines with linguistic interpretation of the kernel matrix

Jacek M. Łęski

Institute of Electronics, Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland

Received 24 May 2004; received in revised form 5 August 2005; accepted 28 September 2005
Available online 25 October 2005

Abstract

Initially, the idea of approximate reasoning using generalized modus ponens and a fuzzy implication is recalled. Next, a fuzzy system based on logical interpretation of if–then rules and with parametric conclusions is presented. Then, it is shown that global and local ε-insensitive learning of the above fuzzy system may be presented as the learning of a support vector regression machine with a special type of a kernel matrix obtained from clustering. The kernel matrix may be interpreted in terms of linguistic values based on the premises of if–then rules. A new method of obtaining a fuzzy system by means of a support vector machine (SVM) with a data-dependent kernel matrix is introduced. This paper contains examples of a SVM used to design fuzzy models of real-life data. Simulation results show an improvement in the generalization ability of a fuzzy system learned by the new method compared with traditional learning methods.

Keywords: Support vector machine; Kernel matrix; Fuzzy system; Generalization ability

1. Introduction

The support vector machine (SVM) is an example of kernel-based methods, which work by transforming data into a high-dimensional feature space and searching for linear relations in such a space. The SVM is historically the first kernel-based method proposed by Boser et al. [6]. This approach is based on the main result of the statistical learning theory, i.e. the generalization ability of a machine depends on both the empirical risk on a training set and the complexity of this machine. For an in-depth study of the statistical learning theory see [43–45].

The SVM can be successfully applied to a wide variety of classification (support vector classifier—SVC) and regression (support vector regression—SVR) problems. After the success of the SVM, other kernel-based methods have been introduced: kernel principal component analysis [39,42], kernel Fisher discriminant [34,3], kernel clustering [4,17], kernel blind source separation [33] and kernel independent component analysis [2]. Kernel-based methods have recently played an important role in many engineering fields, such as pattern recognition, approximation, modeling, character recognition, data mining and others.

The performance of every kernel-based method depends on the kernel type selected. However, there are no general theories for choosing a kernel in a data-dependent way. An information-geometric method of modifying the kernel to improve the performance of SVC is introduced by Amari and Wu [1]. A method of learning the kernel in a data-dependent

E-mail address: JL@boss.iele.polsl.gliwice.pl.

0165-0114/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.fss.2005.09.012

way by semidefinite programming is shown in [23]. However, the main disadvantage of kernel-based methods cannot be eliminated: a feature in a high-dimensional space has no physical interpretation acceptable to a human expert.

For the last few years, there has been an increasing interest in fuzzy systems which incorporate tools well-known from the statistical learning theory. Fuzzy clustering with a weighted (or fuzzy) ε-insensitive loss function is introduced in [26,29,31]. The above method leads to improved robustness of outliers with respect to the traditional fuzzy clustering methods. The support vector fuzzy regression machines have been introduced in [19]. The support vector interval regression network has been established in [21]. A differentiable approximation of the misclassification rate with the use of the empirical risk minimization principle to improve learning of a neuro-fuzzy classifier is proposed in [8]. Work [13] describes the support vector fuzzy clustering method. An ε-insensitive approach to learning of neuro-fuzzy systems has been introduced in [27] and extended in [28,30]. A similar approach to learning a classifier, called a fuzzy SVM, has been independently introduced in [24]. The concept of a fuzzy kernel perceptron is presented in [9]. A different approach based on a SVM to learning framework for fuzzy systems has been proposed in [10,11,14]. In this approach the rulebase of a fuzzy system is extracted from learned SVM. However, in this case, a zeroth-order Takagi–Sugeno–Kang fuzzy system is obtained. Membership functions of fuzzy sets based on the premises of if–then rules are assumed to be Gaussian with centers obtained from support vectors and predefined dispersion.

From the above-mentioned methods, the ε-insensitive approach to learning of neuro-fuzzy systems is of special interest in this work. This approach is based on the premise that human learning, as well as thinking, is tolerant to imprecision. Instead of the usually used quadratic loss function, an ε-insensitive loss function is used which assumes a zero loss for the difference between a model and the reality less than some pre-set value, denoted ε. If this difference is greater than ε, then the loss increases linearly. The ε-insensitive learning is based on the connection between fuzzy modeling and the statistical learning theory where easy control of system complexity is possible. Learning tolerant to imprecision always leads to a better generalization ability and robustness to outliers compared with traditional methods [30,32].

The main goal of this work is to show that global and local ε-insensitive learning of a fuzzy system based on logical interpretation of if–then rules and with parametric conclusions may be presented as learning of a SVR machine with a special type of kernel matrix obtained from clustering. The above kernel matrix may be interpreted in terms of linguistic values based on the premises of if–then rules. A new method of obtaining a fuzzy system by means of a SVM with a data-dependent kernel matrix is introduced. The next goal is to investigate the generalization ability of the fuzzy system obtained by means of the SVR machine with a special type of kernel matrix with linguistic interpretation for real-world benchmark data.

This paper is organized as follows: Section 2 presents the introduction of an ε-insensitive learning method of fuzzy systems based on a logical interpretation of if–then rules and with parametric conclusions and Section 3 shows that this approach leads to a quadratic programming problem; Section 4 shows that global and local ε-insensitive learning of a fuzzy system based on logical interpretation of if–then rules and with parametric conclusions may be presented as learning of a SVR machine with a special type of kernel matrix with linguistic interpretation; Section 5 presents simulation results and a discussion of the fuzzy modeling of real-world benchmark data; and finally, conclusions are drawn in Section 6.

2. Fuzzy systems with a logical interpretation of if–then rules

Fuzzy if–then rules are used to capture the human ability to make a decision in an uncertain and imprecise environment. In this section, fuzzy rules with a logical interpretation and with parametric conclusions will be used to recall the important fuzzy systems which are fundamental in further considerations. The reasons for selection of the above systems are twofold: firstly, the fuzzy systems well known from the literature may be treated as a special type; secondly, a wide class of linguistic interpretations of kernel matrix will be obtained.

Let us assume that I fuzzy if–then rules with t-input and one-output (MISO) are given. The ith rule in which the consequent is represented by a fuzzy set $B^{(i)}(\boldsymbol{\theta})$ whose membership function depends on parameter vector $\boldsymbol{\theta}$ may be written in the following form [15]:

$$\Re^{(i)} : \text{IF } X_1 \text{ IS } A_1^{(i)} \text{ AND } \cdots \text{ AND } X_t \text{ IS } A_t^{(i)}, \text{ THEN } Y \text{ IS } B^{(i)}(\boldsymbol{\theta}), \tag{1}$$