Engineering Applications of Artificial Intelligence 26 (2013) 724–738

Contents lists available at SciVerse ScienceDirect

Engineering Applications of Artificial Intelligence

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

Kernel based nonlinear fuzzy regression model ☆

Zhi-gang Su *, Pei-hong Wang, Zhao-long Song

Key Laboratory of Energy Thermal Conversion and Control of Ministry of Education, School of Energy and Environment, Southeast University, Nanjing 210096, China

ARTICLE INFO

Article history:
Received 6 January 2012
Received in revised form
19 March 2012
Accepted 15 May 2012
Available online 20 June 2012

Keywords:
Nonlinear fuzzy regression
Kernel
Fuzzy EM algorithm
Maximum likelihood estimation
Unmeasured parameter
Power plant

ABSTRACT

Recent years have seen a surge of interest in extending statistical regression to fuzzy data. Most of the recent fuzzy regression models have undesirable performance when functional relationships are nonlinear. In this study, we propose a novel version of fuzzy regression model, called kernel based nonlinear fuzzy regression model, which deals with crisp inputs and fuzzy output, by introducing the strategy of kernel into fuzzy regression. The kernel based nonlinear fuzzy regression model is identified using fuzzy Expectation Maximization (EM) algorithm based maximum likelihood estimation strategy. Some experiments are designed to show its performance. The experimental results suggest that the proposed model is capable of dealing with the nonlinearity and has high prediction accuracy. Finally, the proposed model is used to monitor unmeasured parameter level of coal powder filling in ball mill in power plant. Driven by running data and expertise, a strategy is first proposed to construct fuzzy outputs, reflecting the possible values taken by the unmeasured parameter. With the engineering application, we then demonstrate the powerful performance of our model.

1. Introduction

Regression analysis is one of the most popular statistical techniques enabling identification of functional relationship between independent and dependent variables, when both the independent and dependent variables are given as real numbers. However, in many real-life situations, we cannot obtain such standard observations and can only have fuzzy data. Therefore, such real-life situations are quite often outside the scope of the classical regression analysis (Bargiela et al., 2007), and the classical regression should be extended to deal with fuzzy data. Recent years emerge a surge of literatures about real applications of various artificial intelligence techniques in different engineering fields (without claiming of completeness, see Refs. (Chau, 2006; Muttil and Chau, 2007; Cheng et al., 2002; Lin et al., 2006; Jia et al., 2008; Xie et al., 2006). In this paper, we focus on the fuzzy regression topic as well as its applications.

Originally, "classical" statisticians intuitively consider the fuzzy data as ordinal data in an artificial way, e.g., by "1, 2, 3, …" for a rating with respect to the precise scale. Then the classical statistical regression can be directly applied. Such intuition may induce some disadvantages, as remarked by Nather (2006). In real sense, the fuzzy regression model was first introduced by Tanaka et al. (1980, 1982). In recent years, there is a growing literature that formalizes the linear regression model in a fuzzy domain, in which model parameters and/or data are fuzzy, or imprecise, or vague (without claiming of completeness, for instance, see Refs. Bargiela et al., 2007; Celmins, 1987a, 1987b; Chang, 2001; Chang and Ayyub, 2001; Chang and Lee, 1996; Diamond and Tanaka, 1998; D'Urso, 2003; D'Urso et al., 2011; Kacprzyk and Fedrizzi, 1992; Kim and Bishu, 1998; Korner and Nather, 1998; Lee and Chen, 2001; Nather, 1994, 2006; Sakawa and Yano, 1992; Wang and Tsaur, 2000).

There are two main approaches to fuzzy regression analysis. The first one is the possibilistic approach introduced by Tanaka et al. (1980, 1982). According to this approach, fuzzy regression coefficients are estimated minimizing the fuzziness of the estimated response variable, conditional to the constraint that each estimated value lies within a given interval defined on the observed values of the response variable. The second one is the least-square (LS) approach (Celmins, 1987a, 1987b; Diamond, 1988; D'Urso et al., 2011), which extends the LS criterion to the fuzzy setting. This estimation procedure consists in finding the linear model which best approximates the observed data in a given metric space, taking into account the fuzziness of the data. For the interested reader, see Refs. (Celmins, 1987a, 1987b; Chang, 2001; Chang and Ayyub, 2001; Diamond, 1988; Ming et al., 1997; D'Urso and Gastaldi, 2000; D'Urso, 2003; D'Urso et al., 2011).

In the above mentioned fuzzy regression models, the fuzzy data are viewed implicitly or definitely as realizations of fuzzy random variables. In the literature, there have been proposed two

☆ This work is supported by the National Natural Science Foundation of China (No. 51106025) and Jiangsu Science and Technology Support Program (industrial) funded project (No. BE2010183).

* Corresponding author. Tel.: +86 25 83794520; fax: +86 25 57714489.
E-mail address: suzhigang2008@163.com (Z.-g. Su).

0952-1976/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.engappai.2012.05.009

views of fuzzy data (Gebhardt et al., 1998). The first one assumes the data to be intrinsically fuzzy and uses the mathematical formalism of fuzzy random variables. The second one is based on an epistemic interpretation of fuzzy data, which are assumed to "imperfectly specify a value that is existing and precise, but not measurable with exactitude under the given observation conditions". In the second viewpoint, a fuzzy datum is thus seen as a possibility distribution associated to a precise realization of a random variable that has been only partially observed. According to such viewpoint, Denoeux (2011) proposes a new fuzzy regression model by using fuzzy EM algorithm. Denoeux's model provides a new and well motivated solution with very simple interpretation. As will be shown, however, this model cannot deal with the problem where the nonlinearity exists.

More recently, several issues have been addressed to the fuzzy regression (Bargiela et al., 2007; Bisserier et al., 2010; Chan et al., 2010; Kim et al., 2008; Nasrabadi and Hashemi, 2008), one of which is the nonlinear problem. All the aforementioned approaches are naturally linear or too complex to be used, and thus cannot be used to or cannot be easily used to deal with nonlinear problems. In practice real word systems are nonlinear and complex. Therefore, it is desirable to propose a nonlinear regression model dealing with fuzzy data. Nasrabadi and Hashemi (2008) suggest a robust fuzzy regression extending the risk-neutral model proposed by Modarres et al. (2004, 2005) to a nonlinear fuzzy regression model using multilayered feed forward neural networks where weights, biases, input and output variables are assumed to be LR (Left and Right) type fuzzy. In practical engineering, some process parameters, taken as the independent variables (i.e., inputs), can usually be precisely observed, and only the process parameter, taken as dependent variable (i.e., output), can be imprecisely observed or monitored. In such an example, the unmeasured parameter is defined in Refs. (Su and Wang, 2009; Su et al., 2010). In addition, the model with precise structure parameters is efficient and easy to be realized in the view point of practice engineering. Therefore, Nasrabadi and Hashemi's model is inefficient in such case.

With the above short review, we can see that there is no simple way used to deal with nonlinear fuzzy regression problem where inputs and model structure parameters are crisp but output is fuzzy. In this study, we focus on such topic. The main contribution of this study is that, a novel version of nonlinear fuzzy regression model, called kernel based nonlinear fuzzy regression model, is proposed by introducing the strategy of kernel into Denoeux's model. This model can deal with regression problems where nonlinearity exists and it has high prediction accuracy and is easily to be implemented in practice.

The rest of paper is organized as follows. Section 2 briefly introduces the maximum likelihood estimation strategy based on fuzzy EM algorithm. Section 3 proposes the kernel based nonlinear fuzzy regression model with two numerical experiments. Section 4 presents a practical engineering application of the proposed nonlinear fuzzy regression model in power plant. Section 5 concludes the paper.

2. Fuzzy EM algorithm based maximum likelihood estimation strategy

In this section, we briefly recall the preliminary foundations for the consequent work in this paper. The interested reader may refer to article (Denoeux, 2011) for a thorough treatment on the subject. To format the expressions in sequel, we admit the following nomenclature: capital letter in bold denotes matrix, lowercase letter in bold denotes column vector, italic letter denotes scalar variable, and letter with hat "$\sim$" denotes fuzzy number. Without confusion, the italic capital letter is used to indicate random vector variable.

Let $\boldsymbol{X}$, referred to as the complete-data vector, be a random vector, taking value in sample space $\mathcal{X}$ and describing the result of a random experiment. The probability density function (p.d.f.) of $\boldsymbol{X}$ is denoted by $g(\mathbf{x}, \boldsymbol{\psi})$, where $\boldsymbol{\psi}=(\psi_1, \psi_2, \ldots, \psi_d)^{\mathrm{T}}$ is a column vector of unknown parameters with parameter space Ω, where superscript "T" indicates transposition.

If $\mathbf{x}$, a realization of $\boldsymbol{X}$, is known exactly, we could compute the maximum likelihood estimate (MLE) of $\boldsymbol{\psi}$ as any value maximizing the complete-data likelihood function:

$$L(\boldsymbol{\psi};\mathbf{x}) = g(\mathbf{x};\boldsymbol{\psi}) \tag{1}$$

However, $\mathbf{x}$ is usually not observed precisely, e.g., only partial information about $\mathbf{x}$ is available in the form of a fuzzy subset $\tilde{\mathrm{x}}$ of $\mathcal{X}$. Therefore, the complete-data likelihood function (1) should be extended. Given $\tilde{\mathrm{x}}$ and assume its membership function to be the Borel measurable, the probability of fuzzy set $\tilde{\mathrm{x}}$ can be computed according to Zadeh's definition of the probability of a fuzzy event (Zadeh, 1968). Thus, the observed-data likelihood in the imprecise setting can then be defined as.

$$L(\boldsymbol{\psi};\tilde{\mathbf{x}}) = P(\tilde{\mathbf{x}};\boldsymbol{\psi}) = \int_{\chi} \mu_{\tilde{\mathrm{x}}}(\mathrm{x}) g(\mathrm{x};\boldsymbol{\psi}) d\mathrm{x} \tag{2}$$

In the special case where the complete data $\mathbf{x}=(x_1, x_2, \ldots, x_n)^{\mathrm{T}}$ is a realization of an independent identically distributed (i.i.d.) random vector $\boldsymbol{X}=(X_1, X_2, \ldots, X_n)^{\mathrm{T}}$, and assuming the joint membership function $\mu_{\tilde{\mathrm{x}}}$ to be decomposed in the product of $\mu_{\tilde{x}_i} (i=1, 2, \ldots, n)$, i.e.,

$$\mu_{\tilde{\mathrm{x}}}(\mathbf{x}) = \prod_{i=1}^{n} \mu_{\tilde{x}_i}(x_i) \tag{3}$$

the likelihood function (2) can be written as a product of n items:

$$L(\boldsymbol{\psi};\tilde{\mathbf{x}}) = \prod_{i=1}^{n} \int \mu_{\tilde{x}_i}(x) g(x;\boldsymbol{\psi}) dx \tag{4}$$

and the observed-data log likelihood is:

$$logL(\boldsymbol{\psi};\tilde{\mathbf{x}}) = \sum_{i=1}^{n} \log \int \mu_{\tilde{\mathbf{x}}_i}(x) g(x;\boldsymbol{\psi}) dx \tag{5}$$

The fuzzy EM algorithm approaches the problem of maximizing the observed-data log likelihood $\log L(\boldsymbol{\psi},\tilde{\mathbf{x}})$ by proceeding iteratively with the complete-data likelihood $\log L(\boldsymbol{\psi}, \mathbf{x}) = \log g(\mathbf{x}, \boldsymbol{\psi})$. Each iteration of the fuzzy EM algorithm involves two steps called the expectation step (E-step) and the maximization step (M-step).

The E-step consists in the calculation of

$$Q(\boldsymbol{\psi},\boldsymbol{\psi}^{(q)}) = \mathrm{E}_{\boldsymbol{\psi}^{(q)}}(\log\left[L(\boldsymbol{\psi};\mathbf{x})\right] | \tilde{\mathbf{x}}) = \frac{\int \mu_{\tilde{\mathrm{x}}} \log[L(\boldsymbol{\psi};\mathrm{x})] g(\mathrm{x},\boldsymbol{\psi}^{(q)}) d\mathrm{x}}{L(\boldsymbol{\psi}^{(q)};\tilde{\mathrm{x}})} \tag{6}$$

where the expectation of $\log L(\boldsymbol{\psi}, \boldsymbol{X})$ is taken with respect to the conditional p.d.f. of $\mathbf{x}$ given $\tilde{\mathbf{x}}$, using parameter vector $\boldsymbol{\psi}^{(q)}$:

$$g(\mathbf{x}|\tilde{\mathbf{x}};\boldsymbol{\psi}^{(q)}) = \frac{\mu_{\tilde{\mathrm{x}}}(\mathbf{x}) g(\mathbf{x}|\boldsymbol{\psi}^{(q)})}{\int \mu_{\tilde{\mathrm{x}}}(\mathbf{u}) g(\mathbf{u}|\boldsymbol{\psi}^{(q)}) d\mathbf{u}}$$

The M-step requires the maximization of $Q(\boldsymbol{\psi}, \boldsymbol{\psi}^{(q)})$ with respect to $\boldsymbol{\psi}$ over the parameter space Ω, i.e., finding such that

$$Q(\boldsymbol{\psi}^{(q+1)},\boldsymbol{\psi}^{(q)}) \geq Q(\boldsymbol{\psi},\boldsymbol{\psi}^{(q)}), \quad \boldsymbol{\psi} \in \Omega$$

The fuzzy EM algorithm alternately repeats the E- and M-steps above until the increase of observed-data likelihood becomes smaller than some threshold.