Fuzzy Sets and Systems 122 (2001) 513–526

www.elsevier.com/locate/fss

Fuzzy linear regression analysis for fuzzy input–output data using shape-preserving operations

Dug Hun Hong[a,*], Sungho Lee[b], Hae Young Do[c]

[a] School of Mechanical and Automotive Engineering, Catholic University of Taegu-Hyosung, Kyongsan-si, 330 Kumark 1-ri Hayang-up, Kyungbuk, 712-702, South Korea

[b] Department of Statistics, Taegu University, 712-714, South Korea

[c] Department of Statistics, Kyungpook National University, 702-701, South Korea

Received 5 August 1998; received in revised form 3 September 1999; accepted 20 September 1999

Abstract

In this paper, we have presented a new method to evaluate fuzzy linear regression models based on Tanaka's approach, where both input data and output data are fuzzy numbers, using T_w-based fuzzy arithmetic operations. This method simplifies the computation of fuzzy arithmetic operations. General linear program is applied to derive the solutions. We also prove scale-independent property of our models and discuss the effects of outliers.

Keywords: Fuzzy linear regression; Fuzzy input–output data; Shape-preserving operations

1. Introduction

Fuzzy linear regression provides a means for tackling regression problems lacking a significant amount of data for determining regression models and with vague relationships between the dependent variable and independent variables.

The concept of fuzzy regression analysis was introduced by Tanaka et al. [16] in 1982, where an LP-based method with symmetric triangular fuzzy parameters was proposed.

The method is recommended for practical situations where decisions often have to be made on the basis of imprecise and partially available data where human estimation is influential. This first attempt of applying fuzzy regression was done using non-fuzzy input experimental data. An extension of the idea was reported by Tanaka et al. [16] comparing the capability to process fuzzy input experimental data. Unfortunately, this model was shown to be scale dependent, a fatal technical flaw for regression analysis, by Redden and Woodall [11]. Heshmaty and Kandel [5] applied this method to forecasting in uncertain environment and Watada [18]

* Corresponding author. Tel.: +82-53-850-2712; fax: +82-53-850-2704.
E-mail address: dhhong@cuth.cataegu.ac.kr (D.H. Hong).

PII: S0165-0114(00)00003-8

applied the idea of fuzzy regression to fuzzy time-series. Fuzzy data analysis, regarded as a non-statistical procedure for possibilistic systems, was reported by Tanaka [15], and Tanaka et al. [17]. Fuzzy regression has also been investigated from the viewpoint of the least-squares regression. Celminš [1,2] and Diamond [3] developed several models for fuzzy least-squares fitting. A collection of recent papers dealing with several approach to fuzzy regression analysis can be found in [7].

In the general fuzzy regression model the input data and the output data are fuzzy, the relationship between the input and output data is given by a fuzzy function, and the distribution of the data is possibilistic. In most papers, however, it should be stressed that while output data is assumed to be a fuzzy number, input data is not a fuzzy number. Recently, Sakawa and Yano [14] introduced fuzzy linear models, where both input data and output data are fuzzy numbers, by using three indices for equalities between fuzzy numbers as a more generalized version of Sakawa and Yano [12,13].

In this paper we concentrate on the model of Tanaka. We introduce shape-preserving fuzzy arithmetic operations based on the sup-t-norm convolution and introduce a new class of fuzzy linear regression models based on Tanaka's approach where both input data and output data are fuzzy numbers applying the operations. This approach helps to simplify the evaluation of fuzzy linear regression whose both coefficients and input data are fuzzy numbers. Then an LP-based method to derive the satisfying solution of the decision making is developed. We also prove scale-independent property of our models and discuss the effects of outliers.

2. Preliminaries

A fuzzy number is a convex subset of the real line R with a normalized membership function.

A triangular fuzzy number $\tilde{a}$ denoted by (a,α,β) is defined as

$$\tilde{a}(t)=\begin{cases}1-\frac{|a-t|}{\alpha} & \text{if } a-\alpha\leqslant t\leqslant a,\\ 1-\frac{|a-t|}{\beta} & \text{if } a\leqslant t\leqslant a+\beta,\\ 0 & \text{otherwise,}\end{cases}$$

where $a\in R$ is the center and $\alpha>0$ is the left spread, $\beta>0$ is the right spread of $\tilde{a}$.

If $\alpha=\beta$, then the triangular fuzzy number is called a symmetric triangular fuzzy number and denoted by (a,α).

An L–R fuzzy number $\tilde{a}=(a,\alpha,\beta)_{\mathrm{LR}}$ is a function from the reals into the interval $[0,1]$ satisfying

$$\tilde{a}(t)=\begin{cases}R(\frac{t-a}{\beta}) & \text{for } a\leqslant t\leqslant a+\beta,\\ L(\frac{a-t}{\alpha}) & \text{for } a-\alpha\leqslant t\leqslant a,\\ 0 & \text{else,}\end{cases}$$

where L and R are non-decreasing and continuous functions from $[0,1]$ to $[0,1]$ satisfying $L(0)=R(0)=1$ and $L(1)=R(1)=0$. If $L=R$ and $\alpha=\beta$, then the symmetric L–L fuzzy number is denoted $(a,\alpha)_L$.

A binary operation T on the unit interval is said to be triangular norm (t-norm for short) iff T is associative, commutative, non-decreasing and $T(x,1)=x$ for each $x\in[0,1]$. Moreover, every t-norm satisfies the inequality,

$$T_W(a,b)\leqslant T(a,b)\leqslant \min(a,b)=T_M(a,b),$$

where

$$T_W(a,b)=\begin{cases}a & \text{if } b=1,\\ b & \text{if } a=1,\\ 0 & \text{otherwise.}\end{cases}$$

The crucial importance of $\min(a,b)$, $a\cdot b$, $\max(0,a+b-1)$ and $T_W(a,b)$ is emphasized from a mathematical point of view in Ling [9] among others.