G Model
ASOC-3828; No. of Pages 11
ARTICLE IN PRESS
Applied Soft Computing xxx (2016) xxx–xxx
Contents lists available at ScienceDirect
Applied Soft Computing
journal homepage: www.elsevier.com/locate/asoc

Fuzzy least absolute linear regression☆

Wenyi Zeng[a,*], Qilei Feng[b], Junhong Li[a,c]

[a] College of Information Science and Technology, Beijing Normal University, Beijing 100875, PR China
[b] School of Mathematical Sciences, Beijing Normal University, Beijing 100875, PR China
[c] College of Mathematics and Information Science, Hebei Normal University, Shijiazhuang 050024, PR China

ARTICLE INFO

Article history:
Received 22 October 2013
Received in revised form 3 February 2015
Accepted 22 September 2016
Available online xxx

Keywords:
Triangular fuzzy number
Fuzzy least absolute linear regression
Similarity measure
Parameter estimation

ABSTRACT

The distance between triangular fuzzy numbers is an important research topic in many fields. In this paper, we introduce a new distance between triangular fuzzy numbers, merge least absolute deviation method with the new distance and propose fuzzy regression model. We also investigate the properties and model algorithm of fuzzy least absolute linear regression model in detail by transforming this model into linear programming. Further, we use three numerical examples to illustrate our proposed model reasonable and make some comparisons with some existing fuzzy regression models. Finally, we investigate the robust property of our proposed model and apply our model in the missing data set to verify model data.

1. Introduction

Regression analysis is considered as a very useful data analysis tool that has been widely applied in many fields [3,7]. Least square regression (LSR), i.e., least sum of square deviation, has been widely employed in many fields. However, LSR may behave poorly when there exists bad data in some cases. Least absolute regression(LAR), i.e., least sum of absolute deviation, is preferred when prediction errors are generated from fat-tailed or outlier-producing distribution. But there is still some reluctance to adopt LAR for the analysis of large data sets because it is regarded as computationally highly demanding. Therefore, some authors investigated this topic and obtained some meaning conclusions. For example, Narula and Wellington [17] investigated the minimum sum of absolute error regression, Gentle [12] proposed least absolute values estimation, Charnes et al. [5] proved that LAR can be transformed into linear programming (LP). Of course, the number of variances is increasing dramatically when LAR is transformed into LP in large data sets. Thus, some researchers payed a large attention to the development of algorithms of LAR.

Since many real phenomena are uncertain and imprecise, and cannot be accurately described. Therefore, our studies are needed to be conducted on uncertainty in order to model realistic phenomena based on mathematical techniques. There are two types of uncertainty. First, there exists probabilistic uncertainty which is solved overtime or through experimentation. Secondly, there is fuzzy uncertainty that has nothing to do with experimentation or time. Zadeh [27] described this fuzzy uncertainty as ambiguity and vagueness, and introduced the theory of fuzzy sets to build such a system as needed to deal with ambiguous and vague sentences or information. Chiena and Tsai [8] applied fuzzy numbers to evaluate perceived service quality. Tanaka et al. [24] first introduced fuzzy linear regression in 1982 and constructed fuzzy linear regression models with crisp input, fuzzy output and fuzzy parameters. The objective was to minimize the total spread of fuzzy parameters subject to the support of the estimated values to cover the support of the observed values for a certain h-level. Tanaka et al. [23] explained fuzzy uncertainty of dependent variables with the fuzziness of response functions or regression coefficients in the regression model. Redden and Woodall [21] investigated the properties of certain fuzzy linear regression methods. Savic and Pedrycz [22] studied the evaluation of fuzzy linear regression models. Kim and Bishu [15] investigated fuzzy linear regression model.

As yet, most of the existing papers that have been published on the fuzzy regression model have used the least squares method to construct the fuzzy regression model. For example, Diamond [10] introduced the distance between the triangular fuzzy numbers and proposed fuzzy least square regression (FLSR) to determine fuzzy parameters. Xu and Li [26] proposed multidimensional least squares fitting with fuzzy model. Bargiela et al. [1] investigated multiple regression with fuzzy data. Kao and Chyu [14] proposed

☆ This work is supported by grants from the National Natural Science Foundation of China (10971243).
* Corresponding author.
E-mail address: zengwy@bnu.edu.cn (W. Zeng).

http://dx.doi.org/10.1016/j.asoc.2016.09.029
1568-4946/

least squares estimates in fuzzy regression analysis, Pourahmad et al. [20] applied fuzzy logistic regression in clinical vague status. Bisserier et al. [2] investigated linear fuzzy regression using trapezoidal fuzzy intervals. Chang and Lee [4] proposed another approach to construct the fuzzy regression model based on the ranking of fuzzy numbers, and gave its calculating algorithm. Inuiguchit et al. [13] investigated Mean-Absolute-Deviation-Based fuzzy linear regression analysis. However, the least squares method is so sensitive to outliers that it could be greatly affected by a small number of outliers. Since outliers in the response variable represent model failure, there has been an increased interest in robust estimation procedures, which are insensitive to some outliers, applied to the regression analysis. Like to an ordinary regression, we need robust methods in order to estimate the fuzzy regression coefficients, Chen and Hsueh [6] investigated a mathematical programming method for formulation a fuzzy regression model. Nasrabadi et al. [18] proposed a mathematical programming approach to research fuzzy linear regression analysis. Choi and Buckley [9] pointed out that the least absolute deviation estimators were more efficient than the least squares method in ordinary regression models.

In this paper, we apply the least absolute deviation estimators to construct the fuzzy least absolute linear regression model with crisp inputs, fuzzy output and fuzzy parameters, introduce a distance between triangular fuzzy numbers, propose fuzzy least absolute linear regression model, and use the similarity measure of triangular fuzzy numbers to evaluate the fitting of the observed and estimated values. Three examples show that our proposed model is better than the existing fuzzy regression models studied by some authors used the least squares method. Finally, we investigate the robust property of our proposed model and apply our model in the missing data set to verify model data.

The organization of our work is as follows. In Section 2, some basic notions of fuzzy number, similarity measure of fuzzy sets and some existing fuzzy regression models are reviewed. In Section 3, we introduce the distance between triangular fuzzy numbers, propose fuzzy least absolute regression model based on this distance, investigate its properties for the fuzzy least absolute regression and put forward its calculation method. In Section 4, three numerical examples are provided to compare the proposed model with some existing fuzzy regression models, the robust property of our proposed model is also investigated, and is applied in the missing data set. The conclusion is given in the last section.

2. Preliminary

Throughout this paper, we use $X=\{x_1, x_2, \ldots, x_n\}$ to denote the discourse set, R stands for the all real numbers and $\mathcal{F}(R)$ stands for the set of all fuzzy numbers in R. $\mathcal{F}(X)$ and $\mathcal{P}(X)$ stand for the set of all fuzzy sets and crisp sets in X, respectively. A and $A(x)$ express fuzzy set and its membership function in X, respectively, and the operation "c" stands for the complement operation.

Definition 1 *([11])*. A fuzzy number A is a so-called L–R fuzzy number, $A=(c, l, r)_{LR}$, if the corresponding membership function $A(x)$ satisfies for all $x \in R$

$$A(x)=\begin{cases} L\left(\dfrac{c-x}{l}\right) & c-l \le x \le c, \\ R\left(\dfrac{x-c}{r}\right) & c \le x \le c+r, \\ 0 & else, \end{cases} \tag{1}$$

where c, l, r are the center, left spread and right spread of the fuzzy number A, respectively, and L and R are strictly decreasing continuous functions from [0, 1] to [0, 1] such that $L(0)=R(0)=1$ and $L(1)=R(1)=0$. And $L(x)$ and $R(x)$ are called the left and the right shape function, respectively.

In particular, if $l=r=e$, then the L–R fuzzy number A is called L–R symmetrical fuzzy number and denoted by $A=(c, e)_{LR}$.

Definition 2 *([11])*. A fuzzy number A is called a triangular fuzzy number, $A=(c, l, r)$, if the corresponding membership function $A(x)$ satisfies for all $x \in R$

$$A(x)=\begin{cases} \dfrac{x-(c-l)}{l} & c-l \le x \le c, \\ \dfrac{(c+r)-x}{r} & c \le x \le c+r, \\ 0 & else, \end{cases} \tag{2}$$

In particular, if $l=r=e$, then the triangular fuzzy number A is called symmetrical triangular fuzzy number and denoted by $A=(c, e)$.

Known by the extension principle in Zadeh [27], if $A=(c_1, l_1, r_1)$ and $B=(c_2, l_2, r_2)$ are two triangular fuzzy numbers, then we have the algebraic operations of triangular fuzzy numbers in the following.

$$\begin{aligned} -A &= (-c_1, l_1, r_1) \\ A+B &= (c_1+c_2, l_1+l_2, r_1+r_2) \\ A-B &= (c_1-c_2, l_1+l_2, r_1+r_2) \\ k\cdot A &= (kc_1, |k|l_1, |k|r_1),\ k \in R \\ A\times B &= (c_1c_2, |c_1|l_2+|c_2|l_1, |c_1|r_2+|c_2|r_1) \end{aligned} \tag{3}$$

The similarity measure of fuzzy sets is an important topic in the fuzzy set theory. The similarity measure indicates the similar degree of two fuzzy sets. Wang [25] first put forward the concept of fuzzy sets' similarity measure and gave a computation formula. Since then, the similarity measure has been extensively applied in many fields such as fuzzy clustering, image processing, fuzzy reasoning, and fuzzy neural network [25,26].

Definition 3 *([28])*. A real function $S: \mathcal{F}(X)\times\mathcal{F}(X) \to [0, 1]$ is called the similarity measure of fuzzy sets, if S satisfies the following properties:

(S1) $S(A, A^c)=0$ if A is a crisp set;
(S2) $S(A, B)=S(B, A)$;
(S3) $S(A, B)=1$ iff $A=B$;
(S4) For all $A, B, C \in \mathcal{F}(X)$, if $A\subseteq B\subseteq C$, then $S(A, C)\le S(A, B)$, $S(A, C)\le S(B, C)$.

Then, the following formulas are used to calculate the similarity measure of fuzzy sets A and B.

$$S_1(A, B)=1-\frac{1}{n}\sum_{i=1}^{n}|A(x_i)-B(x_i)| \tag{4}$$

$$S_2(A, B)=\left(\bigvee_{x\in X}(A(x)\wedge B(x))\right)\bigwedge\left(\bigwedge_{x\in X}(A(x)\vee B(x))\right)^c \tag{5}$$

For two given triangular fuzzy numbers $A=(c_1, l_1, r_1)$, $B=(c_2, l_2, r_2)$, without the loss of generality, we assume that $c_1\le c_2$, then we have

$$S_2(A, B)=1+\frac{c_1-c_2}{r_1+l_2}$$

Please cite this article in press as: W. Zeng, et al., Fuzzy least absolute linear regression, Appl. Soft Comput. J. (2016), http://dx.doi.org/10.1016/j.asoc.2016.09.029