

# Dependency between degree of fit and input noise in fuzzy linear regression using non-symmetric fuzzy triangular coefficients

Hong-Wei Ge\*, Shi-Tong Wang

*School of Information, Southern Yangtze University, Wuxi, Jiangsu, People's Republic of China*

Received 2 May 2006; received in revised form 31 January 2007; accepted 14 March 2007

Available online 23 March 2007

## Abstract

Fuzzy linear regression (FLR) model can be thought of as a fuzzy variation of classical linear regression model. It has been widely studied and applied in diverse fields. When noise exists in data, it is a very meaningful topic to reveal the dependency between the parameter  $h$  (i.e. the threshold value used to measure degree of fit) in FLR model and the input noise. In this paper, the FLR model is first extended to its regularized version, i.e. regularized fuzzy linear regression (RFLR) model, so as to enhance its generalization capability; then RFLR model is explained as the corresponding equivalent *maximum a posteriori* (MAP) problem; finally, the general dependency relationship that the parameter  $h$  with noisy input should follow is derived. Particular attention is paid to the regression model using non-symmetric fuzzy triangular coefficients. It turns out that with the existence of typical Gaussian noisy input, the parameter  $h$  is inversely proportional to the input noise. Our experimental results here also confirm this theoretical claim. Obviously, this theoretical result will be helpful to make a good choice for the parameter  $h$ , and to apply FLR techniques effectively in practical applications.

© 2007 Elsevier B.V. All rights reserved.

*Keywords:* Fuzzy linear regression model; Fuzzy regression analysis; Possibility theory; MAP; Optimal parameter choice

## 1. Introduction

Fuzzy linear regression (FLR) provides a means for tackling regression problems lacked of a significant amount of data to determine regression models and with vague relationships between the dependent variable and independent variables. Since the concept of FLR was first introduced by Tanaka et al. [15], the literature dealing with FLR has grown rapidly. For example, a modified version of Tanaka's fuzzy regression model was given in [8], where fuzzy regression for fuzzy input–output data was considered. In [5], fuzzy linear programming was introduced into the modified Tanaka's model. The important properties of fuzzy regression have been studied in [9,10,6]. More variants of Tanaka's model can be seen in [2,7,11–14]. Especially in [2], the support vector technique is introduced into fuzzy regression analysis to enhance its generalization capability. However, how to choose the parameter  $h$  (i.e. the threshold value used to measure degree of fit) for FLR model with noisy input still keeps an open problem. In practice, the input data often contain noise. With the existence of noisy input, one interesting and challenging issue is how to determine the parameter  $h$  in

\* Corresponding author. Tel.: +86 510 88987789.

E-mail address: [ghw8601@163.com](mailto:ghw8601@163.com) (H.-W. Ge).

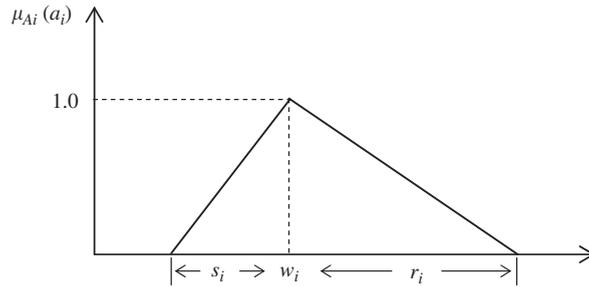


Fig. 1. Non-symmetric triangular fuzzy coefficients.

FLR model. A bad choice for  $h$  will heavily deteriorate the performance of FLR model. Therefore, in this paper, we will pay attention to deriving the dependency between the threshold  $h$  and the input noise.

In recent years, how to choose the loss function and the corresponding parameters for support vector regression (SVR) machine with noisy input has been studied well. Gao and Gunn pointed out that SVR problem could be transformed into the equivalent maximum a posteriori (MAP) problem [1]. Based on this idea, Kwok derived the linear dependency between  $\varepsilon$  and the input noise in  $\varepsilon$ -SVR [3], and Wang et al. investigated the theoretically optimal parameter choices for Huber-SVR with the Huber loss functions and r-SVR with the norm-r loss functions [16]. In this paper, based on the same idea, we first extend FLR model using symmetric and non-symmetric fuzzy triangular coefficients to its regularized version, and then explain this regularized model by using MAP framework [1,3,16], and finally study the dependency between  $h$  and the input noise. The rest of this paper is organized as follows. In Section 2, we will introduce the FLR model and the corresponding regularized fuzzy linear regression (RFLR) model and show the equivalent relationship between RFLR model and MAP. The analysis of the inverse linear dependency between  $h$  and the standard deviation of Gaussian noisy input is displayed in Section 3, while in Sections 4 and 5 the experimental results and some concluding remarks are provided, respectively.

## 2. FLR model and MAP

Fuzzy regression analysis using fuzzy linear models with symmetric triangular fuzzy number coefficients has been formulated earlier. Yen et al. extended the results of an FLR model that uses symmetric triangular coefficients to one with non-symmetric fuzzy triangular coefficients successfully [17]. The need for non-symmetric triangular coefficients arises due to the fact that during the regression using symmetric coefficients, the obtained regression line may not be the best-fitting line. This occurs because of the existence of large number of outliers and higher values of residuals. There are data sets that generate scatter plots in which the data do not fall symmetrically on both sides of the regression line.

### 2.1. FLR model

To introduce the nomenclature, the FLR technique [17] is summarized in the following.

Consider the function

$$Y = f(\mathbf{x}, \mathbf{A}) = A_0 + A_1x_1 + A_2x_2 + \dots + A_nx_n, \tag{1}$$

where  $\mathbf{x} = (1, x_1, x_2, \dots, x_n)^T$  is a vector of non-fuzzy inputs, and  $\mathbf{A} = (A_0, A_1, \dots, A_n)$  is a vector of fuzzy model parameters. Parameters  $A_i$  ( $0 \leq i \leq n$ ) are non-symmetric triangular fuzzy coefficients, and they can be described by the triplets  $\{s_i, w_i, r_i\}$ , where  $w_i$  is the point at which  $\mu_{A_i}(a_i) = 1$ ,  $s_i$  is the left-side spread from the peak point  $w_i$ , and  $r_i$  represents the right-side spread as shown in Fig. 1.

Another representation is also possible, if we normalize the spreads. We can use either spread as the base to normalize the other one. Let us choose  $s_i$  as the base, then  $r_i$  can be expressed as  $r_i = k_i s_i$ , where  $k_i$  are the skew factors and are positive real numbers. The selection of the values for  $k_i$  will be based on the knowledge of the problem and data characteristics. Then  $A_i$  can be described by the triplets  $\{s_i, w_i, k_i s_i\}$ . If the values for  $k_i$  ( $0 \leq i \leq n$ ) are all selected to

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات