Insurance: Mathematics and Economics 45 (2009) 236–241

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

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

Neural networks approach for determining total claim amounts in insurance

Turkan Erbay Dalkilic [a], Fatih Tank [b,*], Kamile Sanli Kula [c]

[a] Karadeniz Technical University, Faculty of Arts and Sciences, Department of Statistics and Computer Sciences, 61080, Trabzon, Turkey
[b] Kirikkale University, Faculty of Arts and Sciences, Department of Statistics, 71100 Yahsihan-Kirikkale, Turkey
[c] Ahi Evran University, Faculty of Arts and Sciences, Department of Mathematics, 40200 Kirsehir, Turkey

ARTICLE INFO

Article history:
Received February 2009
Received in revised form June 2009
Accepted 24 June 2009

Keywords:
Neural networks
Least squares method
Total claim amount
Claim amount payments
Fuzzy if-then rules

ABSTRACT

In this study, we present an approach based on neural networks, as an alternative to the ordinary least squares method, to describe the relation between the dependent and independent variables. It has been suggested to construct a model to describe the relation between dependent and independent variables as an alternative to the ordinary least squares method. A new model, which contains the month and number of payments, is proposed based on real data to determine total claim amounts in insurance as an alternative to the model suggested by Rousseeuw et al. (1984) [Rousseeuw, P., Daniels, B., Leroy, A., 1984. Applying robust regression to insurance. Insurance: Math. Econom. 3, 67–72] in view of an insurer.

1. Introduction

An insurance company is supposed to keep itself ready for all uncertain events such as claim amount demands from the insured. Consequently, in order to estimate future payments of claims, the insurance company sets up various models and checks their validity. Multiple linear regression is one of the most widely used statistical tools in practice. In actuarial statistics, situations occur that do not fit comfortably in such settings and they may generate some critical problems due to strong assumptions. Many problems can be prevented by using Generalized Linear Models, investigated by Kaas et al. (2008), instead of ordinary multiple linear regression. Many studies can be found as an alternative to multiple linear regression in the literature.

There are many studies on the use of the neural networks for parameter estimation. A fuzzy adaptive network approach was established for fuzzy regression analysis by Cheng and Lee (1999) and it was studied on both fuzzy adaptive networks and the switching regression model (Cheng and Lee, 2001). Jang (1993) studied adaptive networks based on a fuzzy inference system. In a study of Takagi and Sugeno (1985), the method for identifying a system using its input–output data was presented. James and Donalt (1999) studied fuzzy regression using neural networks.

There are different studies of fuzzy clustering and the validity criterion. In the study of Mu-Song and Wang (1999), the analysis of fuzzy clustering was done for determining fuzzy memberships, and in this study a method was suggested for indicating the optimal cluster numbers that belong to the variables. Xie and Beni (1991) suggested a validity criterion for fuzzy clustering. In this study we used the Xie–Beni validity criterion for determining optimal cluster numbers.

Various studies have used fuzzy clustering in insurance, such as Verrall and Yakoubov (2008), who specified a data-based procedure for grouping by age, using a fuzzy c-means algorithm. Ebanks et al. (1992) presented how to use the measures of fuzziness to risk classification for life insurance. The article by Horgby (1998) describes how to classify by using a fuzzy inference methodology instead of a risk classification according to the numerical rating system. In the study of Shapiro (2004), fuzzy clustering and the other fuzzy logic topics are discussed.

Detailed information about the historical development of neural networks, fuzzy logic and genetic algorithms and their useful application areas in insurance can be found in Shapiro (2002).

In this paper, we intend to highlight the importance of the neural networks approach to estimating total claim amount payments. The remainder of the paper is organized as follows. Section 2 introduces the parameter estimation in multiple linear regression. In Section 3, the fuzzy if-then rules and the use of these rules are introduced, using adaptive networks for analysis. In Section 4, an algorithm for parameter estimation using a neural network is given,

* Corresponding author. Tel.: +90 5322674950.
E-mail address: fatihtank@gmail.com (F. Tank).

0167-6687/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.insmatheco.2009.06.004

and in Section 5 a numerical study, based on real data, is investigated by using the algorithm suggested in Section 4. Finally, in Section 6, a discussion and conclusion are provided.

2. Multiple linear regression analysis

The main aim of regression analysis is to explain a dependent variable through its relationship with p independent variables. Consider a multiple linear regression model with p independent variables:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon. \tag{1}$$

In this sense the insurer wants to find linear relations between the dependent and independent variables in insurance. The observations, $y_1, y_2, \ldots, y_n$, recorded for each of these p levels can be expressed in the following way:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + \cdots + \beta_p x_{p1} + \varepsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_{12} + \beta_2 x_{22} + \cdots + \beta_p x_{p2} + \varepsilon_2 \\ &\vdots \\ y_n &= \beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n} + \cdots + \beta_p x_{pn} + \varepsilon_n. \end{aligned} \tag{2}$$

The system of n equations shown previously can be represented in matrix notation as follows:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{3}$$

The matrix $\mathbf{X}$ in Eq. (3) is referred to as the design matrix and it contains information about the levels of the predictor variables at which the observations are obtained. The vector $\boldsymbol{\beta}$ contains all the regression coefficients. To obtain the regression model, $\boldsymbol{\beta}$ should be known and can be estimated using least squares (LS) estimates. The following equation can be obtained:

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}. \tag{4}$$

Knowing the estimates, $\widehat{\boldsymbol{\beta}}$, the multiple linear regression model can be estimated, and the sum of squares error can be calculated as

$$\widehat{\mathbf{Y}}_{\mathrm{LS}} = \mathbf{X}\widehat{\boldsymbol{\beta}} \tag{5}$$

$$\widehat{\varepsilon}_{\mathrm{LS}} = \sum_{k=1}^{n} (y_k - \widehat{y}_k)^2 \tag{6}$$

respectively.

3. Fuzzy if-then rules and adaptive networks

The adaptive network used in estimating the unknown parameters of the regression model is based on fuzzy if-then rules and a fuzzy inference system. When the problem is to estimate a regression line to fuzzy inputs coming from different distributions, the Sugeno fuzzy inference system is appropriate, and the proposed fuzzy rule in this case is indicated as

$$\begin{aligned} R^K &: \text{IF } \left(x_1 = F_1^K, x_2 = F_2^K, \ldots, x_p = F_p^K\right) \\ &\quad \text{THEN } \left(Y = Y^K = c_0^K + c_1^K x_1 + \cdots + c_p^K x_p\right). \end{aligned} \tag{7}$$

Here F_i^K stands for fuzzy clusters or fuzzy terms associated with the input x_i in the Kth rule, and Y^K is the system output due to rule R^K (Cheng and Lee, 1999; Takagi and Sugeno, 1985). For instance, suppose a data set has two-dimensional input $X = [x_1, x_2]$. For input x_1, there are two fuzzy clusters, "small" and "large", and for input x_2, two fuzzy clusters, "light" and "heavy". In this case the fuzzy inference system contains the following four rules:

$$R^1: \text{IF } (x_1 \text{ is small and } x_2 \text{ is light}) \quad \text{THEN } (Y^1 = c_0^1 + c_1^1 x_1 + c_2^1 x_2) \tag{8}$$

$$R^2: \text{IF } (x_1 \text{ is small and } x_2 \text{ is heavy}) \quad \text{THEN } (Y^2 = c_0^2 + c_1^2 x_1 + c_2^2 x_2) \tag{9}$$

$$R^3: \text{IF } (x_1 \text{ is large and } x_2 \text{ is light}) \quad \text{THEN } (Y^3 = c_0^3 + c_1^3 x_1 + c_2^3 x_2) \tag{10}$$

$$R^4: \text{IF } (x_1 \text{ is large and } x_2 \text{ is heavy}) \quad \text{THEN } (Y^4 = c_0^4 + c_1^4 x_1 + c_2^4 x_2). \tag{11}$$

Here the multivariate input vector $X = [x_1, x_2, \ldots, x_p]$, univariate output vector Y and posterior parameter set c_i^K are crisp numbers, but the degrees of belonging to determined rules of the input vector X are fuzzy. The fuzzy system which is suitable for the fuzzy rules in Eqs. (8)–(11) is illustrated in Fig. 1, and the corresponding equivalent adaptive network architecture is shown in Fig. 2 (Mu-Song and Wang, 1999; Cheng and Lee, 1999; Shapiro, 2002).

There are two levels of nodes in Layer 1. The first level includes nodes "*small*" and "*large*" and the second level includes nodes "*light*" and "*heavy*". The output of the layer is the membership function based on the linguistic value of the input. Nodes in Layer 2 output the products w^L. The function of a node in this layer is to synthesize the information in the premise section of the fuzzy if-then rule; for example, the first node in Layer 2 includes IF (x_1 is small and x_2 is light), and each node output represents the firing strength of a rule. Layer 3 performs a normalization of the output signals from Layer 2. Each node in Layer 4 corresponds to the consequence of each fuzzy if-then rule; for example, the first node in Layer 3 includes THEN ($\widehat{Y}^1 = c_0^1 + c_1^1 x_1 + \cdots + c_p^1 x_p$). Finally, the single node in Layer 5 computes the overall output as the summation of all incoming signals, which is equivalent to performing an aggregation of all the fuzzy if-then rules.

The weighted mean of the models obtained according to fuzzy rules is the output of the Sugeno fuzzy inference system and the common regression model for data coming from different clusters is indicated with this weighted mean.

A neural network enabling the use of a fuzzy inference system for fuzzy regression analysis is known as an adaptive network. Used for obtaining a good approach to regression functions and formed via neurals and connections, such an adaptive network consists of five layers (Hisao and Tanaka, 1992; Hisao and Manabu, 2001; Horia and Costel, 1996).

The nodes which form a network can be separated into two main groups: adaptive nodes and fixed nodes. The nodes which are forming Layer 1 and Layer 4 are called "adaptive nodes", since the nodes located in Layer 1 generate the membership degrees by Eq. (14) which have various values depending on the parameters $\{v_h, \sigma_h\}$. Similarly, the nodes located in Layer 4 generate the outputs, which have various values depending on the parameters c_i^K. In contrast, the nodes which form Layer 2, Layer 3 and Layer 5 do not include any parameter; thus they are depend on the numerical value from the previous layer, and are called "*fixed nodes*".

The fuzzy rule number of the system depends on the number of independent variables and the cluster or fuzzy sets number forming independent variables. There are significant numbers of validity criteria for fuzzy clusters in the literature. In this study, the fuzzy clustering validity function S, also called the Xie–Beni index, which is proposed by Xie and Beni (1991), will be used. This function will be explained in Section 4. When the independent variable number is indicated with p, if the cluster number