USING FUZZY REGRESSION ANALYSIS IN EDUCATIONAL PROCESS

O.M. Poleshuk, E.G. Komarov

Moscow State Forest University,
141005, Mytishchy-5, Moscow area, Russia
olga.pol@bk.ru, polig@infoline.su

The purpose of this paper is to develop a fuzzy regression model for linguistic descriptions of the type: "unsatisfactory", "satisfactory", "good", "excellent", which may be triangular and trapezoidal fuzzy numbers [1]. The developed model is suggested to be used in educational process for students' grades predictions.

Let $\tilde{Y}_i \equiv (y_1^i, y_2^i, y_L^i, y_R^i)$, $\tilde{X}_j^i \equiv (x_1^{ji}, x_2^{ji}, x_L^{ji}, x_R^{ji})$, $j = \overline{1,m}$ $i = \overline{1,n}$ – be output and input trapezoidal fuzzy numbers, $\tilde{a}_j \equiv (b^j, b_L^j, b_R^j)$, $j = \overline{0,m}$ – unknown coefficients, which are defined as triangular numbers. The linear hybrid model relates $\tilde{Y}$ to $\tilde{X}_j$, $j = \overline{1,m}$, as follows:

$$\tilde{Y} = \tilde{a}_0 + \tilde{a}_1\tilde{X}_1 + \ldots + \tilde{a}_m\tilde{X}_m.$$

The model building starts with input and output numbers being exchanged with corresponding weighted intervals. The weighted interval for the fuzzy number $\tilde{A} \equiv (a_1, a_2, a_L, a_R)$ is an interval $[A_1, A_2]$, such as [2]

$$A_1 = a_1 - \frac{1}{6}a_L, \quad A_2 = a_2 + \frac{1}{6}a_R.$$

We shall define the distance between two fuzzy numbers $\tilde{A}$, $\tilde{B}$ with weighted intervals $[A_1, A_2]$, $[B_1, B_2]$ as follows:

$$f(\tilde{A}, \tilde{B}) = \sqrt{(A_1 - B_1)^2 + (A_2 - B_2)^2}.$$

We shall consider the function F that equals to the sum of the squared distances between observed and predicted data over i from 1 to n.

Optimizational problem is the following: $F(b^j, b_L^j, b_R^j) = \sum_{i=1}^n f^2(\hat{Y}_i, \tilde{Y}_i) \to \min$, $b_L^j \geqslant 0$, $b_R^j \geqslant 0$, $j = \overline{0,m}$. The problem can will be solved by the known methods [1].

Data of triangular and trapezoidal fuzzy numbers are used to illustrate the solutions for developed fuzzy regression. These fuzzy numbers are students' grades formalizations [1]. The same set of data was applied to fuzzy regression and ordinary regression. Therefore, comparisons among different results can be made. The hybrid regression and the ordinary least-squares regression are listed below.

$$\tilde{Y} = (0, 0.566, 0) + (0.412, 0.104, 0)\tilde{X}_1 + (0.466, 0, 0)\tilde{X}_2 + (0.130, 0, 0)\tilde{X}_3, \tag{1}$$

$$S_{\tilde{y}} = 0.454, \quad HR = 0.827, \quad HS_e = 0.213,$$

$$Y = 0.708 + 0.301X_1 + 0.428X_2 + 0.394X_3,$$

$$S_Y = 0.949, \quad R = 0.808, \quad S_e = 0.509. \tag{2}$$

It can be observed that HR of fuzzy regression is greater than R of ordinary regression model. Moreover authors believe that the ordinary least-squares regression model is not altogether correct for students' grades as all arithmetic operations are incorrect in the order scale.

The true predictions for fuzzy regression and ordinary regression are the following: 90 % and 60 % . Thus the developed fuzzy regression model can be used in practice with success.

References

1. *Domrachev V.G., Poleshuk O.M.* A regression model for fuzzy initial data // Automation and Remote Control. 2003. V. 64, N. 11. P. 1715–1724.

2. *Poleshuk O.M., Komarov E.G.* New defuzzification method based on weighted intervals // Proceedings of the 27th International Conference of the North American Fuzzy Information Processing Society. NAFIPS'2008, New York, May 19–22, 2008.