УДК 681.3

О ПОСТРОЕНИИ РЕГРЕССИОННОЙ ЗАВИСИМОСТИ ПРИ ИСХОДНЫХ НЕЧЕТКИХ ДАННЫХ T - ТИПА

Домрачев В.Г., Полещук О.М.

Развитие методов нечеткого регрессионного анализа началось с работы /1/, в которой была предложена первая нечеткая линейная регрессионная модель. Эта модель расширила область применения методов классического регрессионного анализа, поскольку позволила строить регрессионные зависимости на основе нечеткой исходной информации. Особая значимость предложенной в /1/ модели состояла в том, что исходная информация могла иметь как количественный, так и качественный характер, что открывало перспективы применения этой модели в ряде областей деятельности человека. Первая нечеткая регрессионная модель вызвала большой интерес, следствием чего стало появление значительного числа работ, в которых на основе различных оптимизационных критериев строились нечеткие регрессионные модели /2-4/ или комбинированные нечеткие регрессионные модели /5-11/, которые сочетали в себе элементы классической регрессионной модели и нечеткой регрессионной модели. Классическая регрессионная модель опирается на аппарат теории вероятностей и поэтому учитывает в модели одну из разновидностей неопределенности – случайность. Нечеткая регрессионная модель опирается на аппарат теории возможностей и теории нечетких множеств /12/ и поэтому учитывает другую разновидность неопределенности – нечеткость. Подобный односторонний подход к учету заложенной в модели неопределенности не устроил ряд исследователей, результатом чего стало появление комбинированных нечетких регрессионных моделей.

В связи с рассмотрением ограниченного класса нечетких исходных данных (как правило, рассматривается класс треугольных чисел /12/) в методах нечеткого регрессионного анализа возник пробел, который частично был восполнен в /11/. В настоящей работы строится комбинированная нечеткая регрессионная модель на основе нечеткой исходной информации, представленной данными T - типа (T - числами).

Нечетким числом $\tilde{A}$ называется нечеткое подмножество множества действительных чисел R, имеющее функцию принадлежности

$$\mu_A(x): R \to [0,1].$$

Нечеткое число $\tilde{A}$ называется нормальным, если $\max\limits_x \mu_A(x) = 1$.

T - числом будем называть нормальное нечеткое число с функцией принадлежности

$$\mu_A(x)=\begin{cases} 1+\dfrac{x-a_1}{a_L},\ x\le a_1,\ a_L>0 \\ 1-\dfrac{x-a_2}{a_R},\ x\ge a_2,\ a_R>0 \\ 1,\ x\in[a_1,a_2] \\ 0,\ x\in(-\infty,a_1-a_L\bigcup[a_2+a_R,+\infty) \end{cases}.$$

Отрезок $[a_1,a_2]$ называется интервалом толерантности, а a_L и a_R - левым и правым коэффициентами нечеткости соответственно. Символически T - число записывается в виде $\widetilde{A}\equiv(a_1,a_{2,}a_L,a_R)$. Нормальное треугольное число является частным случаем T - числа при $a_1=a_2$ и символически записывается в виде $\widetilde{A}\equiv(a_1,a_L,a_R)$. При $a_L=a_R$ треугольное число называется симметричным, а при $a_L\neq a_R$ несимметричным.

Множество
$A_\alpha=\{x\in R:\mu_A(x)\ge\alpha\}=[A_\alpha^1,A_\alpha^2]=[a_1-(1-\alpha)a_L,a_2+(1-\alpha)a_R],\ \alpha\in[0,1]$
называется множеством $\alpha-$уровня T - числа $\widetilde{A}\equiv(a_1,a_{2,}a_L,a_R)$.

В /10/ для нормального треугольного числа $\widetilde{B}\equiv(b,b_L,b_R)$ введено понятие взвешенной точки

$$B=\frac{\int\limits_0^1\left(\dfrac{b-(1-\alpha)b_L+b+(1-\alpha)b_R}{2}\right)\alpha d\alpha}{\int\limits_0^1\alpha d\alpha}=b+\frac{1}{6}(b_R-b_L).$$

Используя это понятие, определим понятие взвешенного отрезка для T - числа.

Определение 1. Взвешенным отрезком T - числа называется объединение взвешенных точек всех нормальных треугольных чисел, принадлежащих этому числу.

Замечание 1. Определения в работе нумеруются, если понятие вводится впервые.

Замечание 2. Поскольку нормальное треугольное число является частным случаем T - числа, то определение взвешенного отрезка справедливо и для него. Необходимо отметить, что для целого класса симметричных треугольных чисел $\widetilde{A}\equiv(a_1,x,x),x>0,x\in R$ взвешенная точка определяется равной a_1 независимо от коэффициента нечеткости. Определение взвешенного отрезка позволяет учесть степень нечеткости (размытости) числа и в зависимости от коэффициента нечеткости поставить каждому числу в соответствие свой взвешенный отрезок.

Утверждение 1. Взвешенным отрезком T - числа $\widetilde{A} \equiv (a_1, a_2, a_L, a_R)$ является отрезок $[A_1, A_2]$, $A_1 = a_1 - \frac{1}{6} a_L$, $A_2 = a_2 + \frac{1}{6} a_R$.

Символически будем писать $\widetilde{A} \cong [A_1, A_2]$.

Доказательство утверждения 1. Рассмотрим два треугольных числа $\widetilde{B}_1 \equiv (a_1, a_L, 0)$, $\widetilde{B}_2 \equiv (a_2, 0, a_R)$, которые принадлежат T числу $\widetilde{A} \equiv (a_1, a_2, a_L, a_R)$. Обозначим множества $\alpha -$ уровня $\widetilde{B}_1$, $\widetilde{B}_2$ соответственно за $B_{1\alpha} = [B^1_{1\alpha}, a_1]$, $B_{2\alpha} = [a_2, B^2_{2\alpha}]$ и определим взвешенные точки для $\widetilde{B}_1$, $\widetilde{B}_2$:

$$A_1 = \frac{\int_0^1 \left(\frac{B^1_{1\alpha} + a_1}{2} \right) \alpha d\alpha}{\int_0^1 \alpha d\alpha} = \frac{\int_0^1 (2a_1 - (1-\alpha) a_L) \alpha d\alpha}{2\int_0^1 \alpha d\alpha} = a_1 - \frac{1}{6} a_L,$$

$$A_2 = \frac{\int_0^1 \left(\frac{a_2 + B^2_{2\alpha}}{2} \right) \alpha d\alpha}{\int_0^1 \alpha d\alpha} = \frac{\int_0^1 (2a_2 + (1-\alpha) a_R) \alpha d\alpha}{2\int_0^1 \alpha d\alpha} = a_2 + \frac{1}{6} a_R.$$

Рассмотрим произвольное нормальное треугольное число $\widetilde{B} \equiv (b, b_L, b_R)$, которое принадлежит T - числу $\widetilde{A} \equiv (a_1, a_2, a_L, a_R)$. $\widetilde{B}$ имеет множество α - уровня $[B^1_\alpha, B^2_\alpha]$ и взвешенную точку B Из определения принадлежности одного нечеткого числа другому /12/ следует, что $B^1_{1\alpha} \le B^1_\alpha$, $a_1 \le B^2_\alpha$, $a_2 \ge B^1_\alpha$, $B^2_{2\alpha} \ge B^2_\alpha$. Следовательно $\frac{B^1_{1\alpha} + a_1}{2} \le \frac{B^1_\alpha + B^2_\alpha}{2}$, $\frac{a_2 + B^2_{2\alpha}}{2} \ge \frac{B^1_\alpha + B^2_\alpha}{2} \Rightarrow A_1 \le B$, $A_2 \ge B$. Утверждение 1 доказано.

Утверждение 2. Взвешенный отрезок суммы T чисел равен сумме взвешенных отрезков этих чисел.

Доказательство утверждения 2. Покажем, что сумма T - чисел $\tilde{A} \equiv (a_1, a_2, a_L, a_R)$, $\tilde{B} \equiv (b_1, b_2, b_L, b_R)$ с взвешенными отрезками соответственно $\tilde{A} \cong [A_1, A_2]$, $\tilde{B} \cong [B_1, B_2]$, имеет взвешенный отрезок $[A_1 + B_1, A_2 + B_2]$.

Из определения операции сложения для T - чисел /12/ следует, что множество α - уровня $\tilde{A} + \tilde{B}$ имеет вид $[a_1 + b_1 - (1-\alpha)(a_L + b_L);\ a_2 + b_2 + (1-\alpha)(a_R + b_R)]$, поэтому координаты взвешенного отрезка находятся следующим образом:

$$C_1 = \int_0^1 (2(a_1 + b_1) - (1-\alpha)(a_L + b_L))\alpha d\alpha = a_1 + b_1 - \frac{1}{6}(a_L + b_L) = A_1 + B_1,$$

$$C_2 = \int_0^1 (2(a_2 + b_2) + (1-\alpha)(a_R + b_R))\alpha d\alpha = a_2 + b_2 + \frac{1}{6}(a_R + b_R) = A_2 + B_2.$$

Таким образом, $\tilde{A} + \tilde{B} \cong [C_1, C_2] = [A_1 + B_1, A_2 + B_2]$. Утверждение 2 доказано.

Утверждение 3. Пусть T число $\tilde{A} \equiv (a_1, a_2, a_L, a_R)$ имеет взвешенный отрезок $\left[a_1 - \frac{1}{6}a_L, a_2 + \frac{1}{6}a_R\right]$, тогда T - число, полученное умножением $\tilde{A}$ на четкое число a имеет взвешенный отрезок $\left[a\left(a_1 - \frac{1}{6}a_L\right), a\left(a_2 + \frac{1}{6}a_R\right)\right]$ при $a \geq 0$ и $\left[a\left(a_2 + \frac{1}{6}a_R\right), a\left(a_1 - \frac{1}{6}a_L\right)\right]$ при $a < 0$.

Доказательство утверждения 3. Пусть $a \geq 0$. По правилу умножения T - числа $\tilde{A} \equiv (a_1, a_2, a_L, a_R)$ на неотрицательное четкое число получаем T - число $a\tilde{A}$ с множеством α - уровня $a\tilde{A}_\alpha = [aa_1 - (1-\alpha)aa_L, aa_2 + (1-\alpha)aa_R]$. Определим взвешенный отрезок $[D_1, D_2]$ для T - числа $a\tilde{A}$:

$$D_1 = \int_0^1 (2aa_1 - (1-\alpha)aa_L)\alpha d\alpha = aa_1 - \frac{1}{6}aa_L,$$

$$D_2 = \int_0^1 (2aa_2 + (1-\alpha)aa_R)\alpha d\alpha = aa_2 + \frac{1}{6}aa_R.$$

Пусть $a < 0$. По правилу умножения T числа $\tilde{A} \equiv (a_1, a_2, a_L, a_R)$ на отрицательное четкое число получаем T - число $a\tilde{A}$ с множеством α - уровня $a\tilde{A}_\alpha = [aa_2 + (1-\alpha)aa_R, aa_1 - (1-\alpha)aa_L]$. Определим взвешенный отрезок $[D_1, D_2]$ для T - числа $a\tilde{A}$:

$$D_1 = \int_0^1 (2aa_2 + (1-\alpha)aa_R)\alpha d\alpha = aa_2 + \frac{1}{6}aa_R,$$

$$D_2 = \int_0^1 (2aa_1 - (1-\alpha)aa_L)\alpha d\alpha = aa_1 - \frac{1}{6}aa_L.$$

Утверждение 3 доказано.

Для двух T -чисел $\tilde{A} \equiv (a_1, a_2, a_L, a_R)$, $\tilde{B} \equiv (b_1, b_2, b_L, b_R)$ с взвешенными отрезками $\tilde{A} \cong [A_1, A_2]$, $\tilde{B} \cong [B_1, B_2]$ определим меру близости:

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

Займемся построением комбинированной нечеткой линейной регрессионной модели.

Пусть $\tilde{Y} = \begin{pmatrix} \tilde{Y}_1 \\ \dots \\ \tilde{Y}_n \end{pmatrix}$, $\tilde{Y}_i \equiv (y_1^i, y_2^i, y_L^i, y_R^i)$, $i = \overline{1,n}$ выходные T числа,

$\tilde{X}_j = \begin{pmatrix} \tilde{X}_j^1 \\ \dots \\ \tilde{X}_j^n \end{pmatrix}$, $\tilde{X}_j^i \equiv (x_1^{ji}, x_2^{ji}, x_L^{ji}, x_R^{ji})$, $j = \overline{1,m}$, $i = \overline{1,n}$ - входные T -числа.

Зависимость между входными и выходными данными будем искать в виде:

$$\tilde{Y} = a_0 + a_1\tilde{X}_1 + \dots + a_m\tilde{X}_m,$$

a_j, $j = \overline{0,m}$ - неизвестные коэффициенты регрессионной модели.

Используя утверждение 1, определим взвешенные отрезки для наблюдаемых выходных данных $\tilde{Y}_i \cong [Z_{i1}, Z_{i2}] = \left[y_1^i - \frac{1}{6} y_L^i, y_2^i + \frac{1}{6} y_R^i \right]$.

Используя утверждения 1-3, определим взвешенные отрезки для модельных выходных данных

$$\hat{Y}_i = a_0 + a_1\tilde{X}_1^i + \ldots + a_m\tilde{X}_m^i \cong$$

$$\left[a_0 + \sum_{j:a_j\ge 0} a_j\left(x_1^{ji} - \frac{1}{6}x_L^{ji}\right) + \sum_{j:a_j<0} a_j\left(x_2^{ji} + \frac{1}{6}x_R^{ji}\right),\ a_0 + \sum_{j:a_j\ge 0} a_j\left(x_2^{ji} + \frac{1}{6}x_R^{ji}\right) + \sum_{j:a_j<0} a_j\left(x_1^{ji} - \frac{1}{6}x_L^{ji}\right)\right].$$

Оптимизационная задача ставится следующим образом:

$$F = \sum_{i=1}^{n} f^2\left(\hat{Y}_i, \tilde{Y}_i\right) \to \min\cdot$$

$$F = \sum_{i=1}^{n}\left\{\left[a_0 - y_1^i + \frac{1}{6}y_L^i + \sum_{j:a_j\ge 0} a_j\left(x_1^{ji} - \frac{1}{6}x_L^{ji}\right) + \sum_{j:a_j<0} a_j\left(x_2^{ji} + \frac{1}{6}x_R^{ji}\right)\right]^2 + \right.$$

$$\left. + \left[a_0 + -y_2^i - \frac{1}{6}y_R^i + \sum_{j:a_j\ge 0} a_j\left(x_2^{ji} + \frac{1}{6}x_R^{ji}\right) + \sum_{j:a_j<0} a_j\left(x_1^{ji} - \frac{1}{6}x_L^{ji}\right)\right]^2\right\} \to \min.$$

Неизвестные коэффициенты $a_j, j = \overline{0,m}$ находятся из системы нормальных уравнений:

$$\frac{\partial F}{\partial a_j} = 0,\ j = \overline{0,m}.$$

Предложенная модель является комбинированной, поскольку сочетает в себе элементы нечеткой и классической регрессионных моделей. Подобное сочетание позволяет определить аналоги стандартного отклонения для наблюдений, коэффициента корреляции и оценки стандартной ошибки.

Определение 2. Аналогом стандартного отклонения для наблюдений назовем величину

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} f^2\left(\tilde{Y}_i, \bar{\tilde{Y}}\right)},$$

$$\bar{\tilde{Y}} \cong \left[\frac{\sum_{i=1}^{n} Z_{i1}}{n}, \frac{\sum_{i=1}^{n} Z_{i2}}{n}\right] = \left[\frac{\sum_{i=1}^{n}\left(y_1^i - \frac{1}{6}y_L^i\right)}{n}, \frac{\sum_{i=1}^{n}\left(y_2^i + \frac{1}{6}y_R^i\right)}{n}\right].$$

Определение 3. Аналогом коэффициента корреляции назовем величину

$$HR = \frac{\sum_{i=1}^{n} f^2\left(\hat{Y}_i, \bar{\tilde{Y}}\right)}{\sum_{i=1}^{n} f^2\left(\tilde{Y}_i, \bar{\tilde{Y}}\right)}.$$

Определение 4. Аналогом оценки стандартной ошибки назовем величину

$$HS = \frac{1}{n-m-1}\sum_{i=1}^{n} f^2\left(\hat{Y}_i, \tilde{Y}_i\right).$$

Пример.

Рассмотрим задачу построения регрессионной зависимости на основе информации образовательного процесса. В таблице 1 представлены данные успеваемости 10 обучающихся по четырем предметам.

Таблица 1.

№ п/п	X_1	X_2	X_3	Y
1	2	3	3	2
2	3	4	3	3
3	3	2	2	2
4	4	4	3	4
5	5	4	5	5
6	4	3	3	3
7	5	5	4	4
8	5	3	4	3
9	4	2	4	3
10	4	4	4	4

Поскольку оценки «2», «3», «4», «5» являются символическим обозначением лингвистических термов «неудовлетворительно», «удовлетворительно», «хорошо», «отлично», то заменим их функциями принадлежности нечетких множеств, соответствующих этим термам. Нечеткие множества являются T - числами и имеют параметры, представленные в таблице 2. Методы построения функций принадлежности терм-множеств лингвистических переменных описаны, например, в /12,13/.

Таблица 2.

№ п/п	$\tilde{X}_1$	$\tilde{X}_2$	$\tilde{X}_3$	$\tilde{Y}$
1	(0; 0.1; 0; 0.1)	(0.15; 0.45; 0.1; 0.3)	(0.35; 0.35; 0.2; 0.3)	(0; 0.1; 0; 0.15)
2	(0.2; 0.4; 0.1; 0.3)	(0.75; 0.85; 0.3; 0.15)	(0.35; 0.35; 0.2; 0.3)	(0.25; 0.6; 0.15; 0.1)
3	(0.2; 0.4; 0.1; 0.3)	(0; 0.05; 0; 0.1)	(0; 0.15; 0; 0.2)	(0; 0.1; 0; 0.15)
4	(0.7; 0.8; 0.3; 0.15)	(0.75; 0.85; 0.3; 0.15)	(0.35; 0.35; 0.2; 0.3)	(0.7; 0.9; 0.1; 0.05)
5	(0.95; 1; 0.15; 0)	(0.75; 0.85; 0.3; 0.15)	(0.9; 1; 0.05; 0)	(0.95; 1; 0.05; 0)
6	(0.7; 0.8; 0.3; 0.15)	(0.15; 0.45; 0.1; 0.3)	(0.35; 0.35; 0.2; 0.3)	(0.25; 0.6; 0.15; 0.1)
7	(0.95; 1; 0.15; 0)	(1; 1; 0.15; 0)	(0.65; 0.85; 0.3; 0.05)	(0.25; 0.6; 0.15; 0.1)
8	(0.95; 1; 0.15; 0)	(0.15; 0.45; 0.1; 0.3)	(0.65; 0.85; 0.3; 0.05)	(0.25; 0.6; 0.15; 0.1)
9	(0.7; 0.8; 0.3; 0.15)	(0; 0.05; 0; 0.1)	(0.65; 0.85; 0.3; 0.05)	(0.25; 0.6; 0.15; 0.1)
10	(0.7; 0.8; 0.3; 0.15)	(0.75; 0.85; 0.3; 0.15)	(0.65; 0.85; 0.3; 0.05)	(0.25; 0.6; 0.15; 0.1)

Построим нечеткую комбинированную линейную регрессионную модель

$$\tilde{Y} = a_0 + a_1\tilde{X}_1 + a_2\tilde{X}_2 + a_3\tilde{X}_3,\ a_j, j = \overline{0,3}.$$

Решение оптимизационной задачи дает значения неизвестных коэффициентов $a_0 = 0,\ a_1 = 0.352,\ a_2 = 0.466,\ a_3 = 0.133$. Модель имеет вид

$$\tilde{Y} = 0.352\tilde{X}_1 + 0.466\tilde{X}_2 + 0.133\tilde{X}_3,$$

$$S = 0.454,\ HR = 0.805,\ HS = 0.2387.$$

Выводы. Построенная в работе комбинированная линейная нечеткая регрессионная модель значительно расширяет область применения методов нечеткого регрессионного анализа, поскольку позволяет строить регрессионные зависимости для нечетких исходных данных как треугольного типа, так и T - типа. Сочетание элементов нечеткой и классической регрессионных моделей дает возможность определения числовых характеристик модели в виде аналогов стандартного отклонения для наблюдений, коэффициента корреляции и оценки стандартной ошибки. Предложенная модель может с успехом применяться для обработки нечеткой информации как количественного, так и качественного характера самых разных областей. Работа выполнена при поддержке гранта РФФИ 01-07-90463.

Список использованных источников

1. Tanaka H., Uejima S., Asai. Linear regression analysis with fuzzy model. IEEE, Systems, Trans. Systems Man Cybernet. SMC-2 (1982) p. 903-907.
2. Chang P.-T., Lee E.S. Fuzzy linear regression with spreads unrestricted in sign // Comput. Math. Appl. 1994. № 28. P. 61-71.
3. Tanaka H., Ishibuchi H. Identification of possibilistic linear models // Fuzzy Sets and Systems. 1991. № 41. 145-160.
4. Tanaka H., Ishibuchi H., Yoshikawa S. Exponential possibility regression analysis // Fuzzy Sets and Systems. 1995. № 69. P. 305-318.
5. Celmins A. Least squares model fitting to fuzzy vector data // Fuzzy Sets and Systems. 1987. № 22. P. 245-269.
6. Celmins A. Multidimensional least-squares model fitting of fuzzy models // Math. Modeling. 1987. № 9. P. 669-690.
7. Chang Y.-H., Ayyub B.M. Reliability analysis in fuzzy regression // Proc. Annual. Conf. of the North American Fuzzy Information Society (NAFIPS 93). Allentown. 1993. P. 93-97.
8. Chang Y.-H., Johnson P., Tokar S., Ayyub B.M. Least- squares in fuzzy regression // Proc. Annual. Conf. of the North American Fuzzy Information Society (NAFIPS 93). Allentown. P. 98-102.
9. Savic D., Pedrycz W. Evaluation of fuzzy regression models // Fuzzy Sets and Systems. 1991. № 39. P. 51-63.
10. Chang Y.-H. Hybrid fuzzy least-squares regression analysis and its reliability measures // Fuzzy Sets and Systems. 2001. № 119. P. 225-246.
11. Таранцев А.А. Принципы построения регрессионных моделей при исходных данных с нечетким описанием // АиТ. 1997. № 11. С. 27-32.
12. Аверкин А.Н., Батыршин И.З., Блишун А.Ф., Силов В.Б., Тарасов В.Б. Нечеткие множества в моделях управления и искусственного интеллекта.- М.: Наука. 1986.
13. Полещук О.М. Некоторые подходы к моделированию системы управления образовательным процессом // Телекоммуникации и информатизация образования. 2002. № 3. С. 54-72.