УДК 004.02 DOI:10.17587/it.26.137-144

Ю. А. Зак, д-р техн. наук, e-mail: yuriy_zack@hotmail.com,
Аахен, Германия

Нечеткий линейный регрессионный анализ, учитывающий характер влияния входных факторов

Рассмотрено решение задач нечеткого регрессионного анализа в условиях, когда входные и выходная переменная представлены нормализованными нечеткими множествами с LR-представлением функции принадлежности самого общего вида, а коэффициенты регрессии — отрицательные или положительные действительные числа. Свободный член уравнения регрессии — нечеткое множество самого общего вида. Предусмотрены ограничения на установленную экспертами степень влияния некоторых входных факторов. Критерий аппроксимации — минимальное абсолютное значение средневзвешенной суммы абсолютных значений координат минимальных и максимальных значений абсцисс λ-сечений функций принадлежности нечетких множеств выходной переменной и ее оценки по fuzzy-регрессионной модели. Коэффициенты регрессии рассчитываются в результате решения некоторого подмножества задач линейного программирования с последующим выбором среди них решения с наилучшим значением критерия оптимальности.

Ключевые слова: *нечеткие регрессионные модели, fuzzy-множества, LR-представление функции принадлежности, сечения функций принадлежности, линейное программирование*

Введение

Классические методы регрессионного анализа работают только со статистическими данными, представленными действительными числами и находят широкое применение при построении математических моделей производственных, технических и экономических систем, а также в макроэкономике, социологии, политологии и медицине (см., например, [1, 2, 4]). Однако в ряде случаев отдельные показатели и параметры могут быть описаны только лингвистическими (например, плохой, удовлетворительный, хороший, превосходный и т. п.) либо булевыми, либо fuzzy-переменными, либо интервальными значениями. В работах [3—5] показано, как такие переменные могут быть представлены нечеткими множествами. В условиях, когда в статистической выборке некоторые или все входные и (или) выходная переменная представлены нечисловой информацией, использование fuzzy-регрессионных моделей является эффективной альтернативой получения количественных зависимостей в случае установленных экспертами качественных закономерностей изучаемых явлений. Результатом расчета на основе таких моделей, как и параметрами модели (т. е. значениями входных и выходных переменных, либо коэффициентов уравнений модели), являются нечеткие множества с функцией принадлежности непрерывного вида. Эти результаты расчета определяют диапазон возможных значений выходной переменной и оценку (некоторый аналог вероятности) получения определяемого значения в пределах данного диапазона. Построение модели представляет собой в данном случае определение оптимальных в некотором смысле коэффициентов модели с учетом нечеткой информации об объекте и субъективных представлений исследователя об оценках адекватности построенной регрессионной модели.

1. Краткий обзор публикаций по построению нечетких регрессионных моделей

Известны три метода fuzzy-регрессионного анализа [4, 6, 7]:

а) нечеткая регрессия, основанная на критерии минимизации нечеткости [6, 7];

б) метод, получивший название FLSRA (fuzzy least-square regression analysis) [6—8], который, в свою очередь, имеет две разновидности, в одной из которых используется критерий максимальной совместимости, а в другой — критерий минимизации нечеткости;

в) регрессия интервала [4].

Все три метода могут в качестве исходной информации использовать как нечеткие множества, представленные функциями принадлежности, так и детерминированные данные.

В расчетах нечетких коэффициентов модели используются два критерия:

а) для всех расчетных данных принадлежность фактического значения выходной переменной к его нечеткой оценке должна быть не ниже некоторого значения, определяемого как уровень доверия;

б) общая нечеткость расчетного значения выходной переменной $\tilde{Y}$ должна быть минимизирована. При этом во многих работах используется дефаззифицированное значение выходной переменной или ее оценки по fuzzy-регрессионной модели, либо критерии вида $\max\{0; \mu_{\bar{Y}_i}[F(\bar{Y}_i)] - \mu_{\bar{Y}_i}(\bar{Y}_i)\}$ — ограниченная разность нечетких чисел, где $\mu_{\bar{Y}_i}(\bar{Y}_i)$ и $\mu_{\bar{Y}_i}[F(\bar{Y}_i)]$ — соответственно функции принадлежности нечеткого множества выходной переменной и ее расчетного значения на основе fuzzy-регрессионной модели. Здесь и в дальнейшем, в отличие от действительных чисел, символами с чертой сверху обозначаются нечеткие множества. Верхний индекс переменных и нечетких множеств определяет индекс соответствующей переменной или коэффициента модели, а нижний — номер комплекта информации. Большинство публикаций по данной тематике либо рассматривали некоторые частные случаи одной из этих общих постановок задачи, либо давали интересные новые приложения ее применения, либо описывали алгоритмы решения известных постановок этой задачи. П. Даймонд [9, 10] ввел новое понятие расстояния на множестве нечетких чисел между прогнозируемыми и экспериментальными данными. В работах Х. Танака (1982 г.) [11, 12], в работе автора [4], в статьях [9, 10, 13] и во многих других публикациях рассмотрена модель линейной регрессии с нечеткими коэффициентами в виде треугольных fuzzy-чисел. В работах [4, 11, 12] для определения значений коэффициентов модели, минимизирующих суммарную средневзвешенную размытость параметров функции принадлежности, рассматриваемую в различных метриках, предложены методы линейного программирования. В 1987 г. А. Селминс [13] и П. Даймонд [9, 10], а также P. Rousseeuw [14], Янг и Лиу в 2003 г. [15] также предложили методику построения моделей нечеткой регрессии методом наименьших квадратов. Эти подходы, комбинированные с методом наименьших квадратов и получившие название FLSRA (fuzzy least-square regression analysis), были предложены Diamond в 1988 г. и Celmiņš в 1987 г. Предложенные методы, в свою очередь, имеют две разновидности, в одной из которых используется критерий максимальной совместимости, а в другой — критерий минимизации квадратичного отклонения.

Для построения критериев аппроксимации использовались различные метрики, среди которых наибольшее распространение получили показатели λ-сечений нечетких множеств (см., например, [14]). В ряде случаев сформулированная оптимизационная задача становится нелинейной и многоэкстремальной. Для решения ее применялись градиентные, поисковые методы и генетические алгоритмы (см., например, [16]).

2. Постановки задачи

Необходимо на основе представительной выборки, заданной матрицей fuzzy-чисел: $|\bar{X}\bar{Y}|$, где $\bar{X}_i^j$, $j = 1, ..., n$, $i = 1, ..., N$, и все выходные переменные $\bar{Y}_i$, $i = 1, ..., N$, — в самом общем случае нечеткие множества, найти нечеткую линейную модель в виде

$$\begin{aligned}\bar{Y} = \bar{A}_1 \otimes \bar{X}_1 + \bar{A}_2 \otimes \bar{X}_2 + \ldots + \\ + \bar{A}_i \otimes \bar{X}_i + \ldots + \bar{A}_n \otimes \bar{X}_n + \bar{A}_0.\end{aligned} \quad (1)$$

Здесь $\bar{A}_1, \bar{A}_2, \ldots, \bar{A}_j, \ldots, \bar{A}_n$ и $\bar{A}_0$ — некоторые fuzzy-множества с заданными (с точностью до неизвестных параметров) функциями принадлежности. Операции "⊗" и "+" — соответственно операции умножения и сложения fuzzy-множеств. Результат вычисления по формуле (1) — также некоторое fuzzy-множество.

Представляет определенный интерес также некоторые частного вида постановки задачи fuzzy-регрессионного анализа:

1) матрица наблюдений $(\bar{X}\bar{Y})$ представлена fuzzy-числами $\bar{X}_i^1, \bar{X}_i^2, \ldots, \bar{X}_i^j, \ldots, \bar{X}_i^n$ и $\bar{Y}_i$, $i = 1, ..., N$, в каждом i-м комплекте информации. Необходимо найти нечеткую линейную регрессионную модель вида

$$\begin{aligned}\bar{Y} = b^1 \times \bar{X}^1 + b^2 \times \bar{X}^2 + \ldots + b^j \times \bar{X}^j + \\ + \ldots + b^n \times \bar{X}^n + \bar{B},\end{aligned} \quad (2)$$

где $\bar{X}_i^1, \bar{X}_i^2, \ldots, \bar{X}_i^j, \ldots, \bar{X}_i^n$ и $\bar{Y}_i$, $j = 1, ..., n$, $i = 1, ..., N$, — некоторые нечеткие множества с заданными функциями принадлежности; коэффициенты линейной модели b^1, b^2, ..., b^j, ..., b^n — некоторые действительные положительные числа, а свободный член уравнения $\bar{B}$ — нечеткое множество;

2) входные переменные $x_i^1, x_i^2, \ldots, x_i^j, \ldots, x_i^n$ — действительные числа, а выходные переменные $\bar{Y}_i$, $i = 1, ..., N$, коэффициенты линейной регрессионной модели и свободный член $\bar{B}^j$, $j = 0, ..., n$ — нечеткие множества.

Отметим, что в частном случае некоторые входные переменные $\bar{X}_{ij}$ и выходная переменная в комплектах информации могут быть представлены действительными числами.

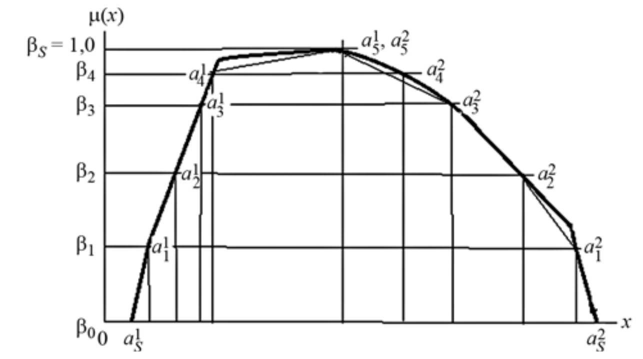
Ниже будет представлено решение задач нечеткого регрессионного анализа в условиях, когда входные и выходная переменная представлены fuzzy-множествами самого общего вида, а коэффициенты регрессии — отрицательные или положительные действительные числа. Свободный член уравнения регрессии — нечеткое множество самого общего вида. Рассмотрены некоторые новые критерии аппроксимации, основанные на сравнении средневзвешенных координат минимальных и максимальных значений абсцисс λ-сечений функций принадлежности нечетких множеств выходной переменной и ее оценки по fuzzy-регрессионной модели, а также координат центров тяжести функций принадлежности fuzzy-множеств. Число сечений, значения функции принадлежности каждого из выбираемых сечений и весовые коэффициенты соответствующих отклонений, которые учитываются в критерии аппроксимации, определяются экспертами и лицом, принимающим решение. Результат расчета fuzzy-регрессионной модели — нечеткое множество, функция принадлежности которого аппроксимируется многоугольником, соединяющим координаты расчетных точек.

В публикациях автора [18, 19] сформулированная задача была решена методом наименьших квадратов в условиях ограничений, что значения коэффициентов регрессии — положительные числа. В данной работе эти ограничения отсутствуют. Кроме того, существенной особенностью рассматриваемой в данной работе постановки задачи является необходимость учета и сохранения в модели качественного влияния отдельных входных факторов (положительного или отрицательного) на выходную переменную, т. е. наличия определенных ограничений. Это не позволяет решить данную задачу методом наименьших квадратов. В этих условиях в данной статье получены детерминированные эквиваленты сформулированных задач и алгоритмы расчета параметров критериев аппроксимации, детерминированных значений коэффициентов уравнения регрессии, а также свободного члена, представленного fuzzy-множеством, определенным с точностью до неизвестных параметров, основанные на решении задач линейного программирования.

Полученные результаты позволяют разработать эффективные алгоритмы построения моделей для многих частных видов функций принадлежности исходных данных задачи (в частности, прямоугольного, треугольного и трапецеидального видов), а также решать многие прикладные проблемы в экономике, логистике, социологии и маркетинге.

3. Критерии аппроксимации, математическая модель и алгоритм решения задачи

Рассматриваются нормализованные нечеткие множества (см. рисунок) с LR-представлением функции принадлежности, т. е. функции, значения функции принадлежности которых $\mu_{\bar{A}}(\bar{A})$ начиная с некоторого значения абс-



Нечеткое множество произвольного вида

цисс $x = a(\bar{A})$, $\mu_{\bar{A}}[a(\bar{A})] = 0$ растут до значения $x = m_1(\bar{A})$, $\mu_{\bar{A}}[m_1(\bar{A})] = 1{,}0$, на некотором отрезке $x \in [m_1(\bar{A}), m_2(\bar{A})]$ (в частном случае, только в одной точке) $x = [m_1(\bar{A}) = m_2(\bar{A}) = m(\bar{A})]$ имеют постоянное значение $\mu_{\bar{A}}[m_1(\bar{A})] = \mu_{\bar{A}}[m_2(\bar{A})] = 1{,}0$, а на отрезке $x \in [m_2(\bar{A}), b(\bar{A})]$ убывают до значения $\mu_{\bar{A}}[b(\bar{A})] = 0$. Среди достаточно большого числа функций этого класса наибольший интерес представляют fuzzy-множества с функцией принадлежности прямоугольного, треугольного и трапециевидного типа.

Обозначим координаты абсцисс крайних точек каждого из этих сечений соответственно $a_k^1[A(\beta_k)]$ и $a_k^2[A(\beta_k)]$, $k = 0, 1, ..., K$, координаты абсцисс крайних точек функций принадлежности соответствующих fuzzy-множеств при $\beta_0 = \beta_K = 0$ — $R^1(\bar{A})$ и $R^2(\bar{A})$, а крайних точек функций принадлежности соответствующих fuzzy-множеств при $\beta_s = 1{,}0$, где $s = \dfrac{K+1}{2}$, — соответственно $T^1(\bar{A})$ и $T^2(\bar{A})$. Здесь, если $T^1(\bar{A}) = T^2(\bar{A})$, т. е. $\mu_{\bar{A}}(A)$ имеет только одну координату абсцисс, для которой $\mu_{\bar{A}}(A) = 1{,}0$, то K — четное число.

Координаты абсцисс соответствующих сечений функций принадлежности определяются по формулам:

$$a_k^1[\bar{A}(\beta_k)] = R^1(\bar{A}) + \beta_k[T^1(\bar{A}) - R^1(\bar{A})];$$
$$a_k^2[\bar{A}(\beta_k)] = R^2(\bar{A}) - \beta_k[R^2(\bar{A}) - T^2(\bar{A})], \quad (3)$$
$$k = 0,\ 1, \ldots, K.$$

В результате умножения нечеткого множества на некоторое действительное число b координаты абсцисс соответствующих сечений функций принадлежности fuzzy-множества $\bar{D}$ вычисляются по формулам:

$$a_k^1[\bar{D}(\beta_k)] = ba_k^1[\bar{A}(\beta_k)] =$$
$$= \begin{cases} b\{R^1(\bar{A}) + \beta_k[T^1(\bar{A}) - R^1(\bar{A})]\}, \text{если } b \geqslant 0; \\ b\{R^2(\bar{A}) - \beta_k[R^2(\bar{A}) - T^2(\bar{A})]\}, \text{если } b < 0; \end{cases} \quad (4)$$

$$a_k^2[\bar{D}(\beta_k)] = ba_k^2[\bar{A}(\beta_k)] =$$
$$= \begin{cases} b\{R^2(\bar{A}) - \beta_k[R^2(\bar{A}) - T^2(\bar{A})]\}, \text{если } b \geqslant 0; \\ b\{R^1(\bar{A}) + \beta_k[T^1(\bar{A}) - R^1(\bar{A})]\}, \text{если } b < 0. \end{cases} \quad (5)$$

В результате сложения fuzzy-множеств координаты соответствующих сечений вычисляются по формулам:

$$a_k^1[\bar{P}(\beta_k)] = \sum_{i=1}^{n} a_k^1[\bar{A}_i(\beta_k)] =$$
$$= \left\{ \sum_{i=1}^{n} R^1(\bar{A}_i) + \sum_{i=1}^{n} \beta_k[T^1(\bar{A}_i) - R^1(\bar{A}_i)] \right\}; \quad (6)$$

$$a_k^2[\bar{P}(\beta_k)] = \sum_{i=1}^{n} a_k^2[\bar{A}_i(\beta_k)] =$$
$$= \left\{ \sum_{i=1}^{n} R^2(\bar{A}_i) - \sum_{i=1}^{n} \beta_k[R^2(\bar{A}_i) - T^2(\bar{A}_i)] \right\}. \quad (7)$$

В задаче fuzzy-регрессионного анализа в каждом из N комплектов информации $i = 1, ..., N$, нечеткие множества входной $\bar{X}_{ik}$ и выходной переменной $\bar{Y}_i$ представляются $2(K + 1)$ детерминированными параметрами, соответственно:

$$a_0^1(\bar{X}_i^j), \ldots, a_k^1(\bar{X}_i^j), \ldots, a_K^1(\bar{X}_i^j), a_0^2(\bar{X}_i^j), \ldots,$$
$$\ldots, a_k^2(\bar{X}_i^j), \ldots, a_K^2(\bar{X}_i^j)$$
$$\text{и } a_0^1(\bar{Y}_i), \ldots, a_k^1(\bar{Y}_i), \ldots, a_K^1(\bar{Y}_i), a_0^2(\bar{Y}_i), \ldots,$$
$$\ldots, a_k^2(\bar{Y}_i), \ldots, a_K^2(\bar{Y}_i).$$

В качестве неизвестных детерминированных параметров (коэффициентов модели) определены детерминированные значения b^1, b^2, ..., b^j, ..., b^n, и уравнение регрессии будем искать в виде

$$a_k^r(\bar{Y}) = \sum_{j=1}^{n} b^j a_k^r(\bar{X}_k^{jr}) + B_k^r, \quad (8)$$
$$r = 1, 2, k = 0, 1, \ \ldots, K.$$

Здесь $\bar{B}$ — нечеткое множество с функцией принадлежности, определенной с точностью до неизвестных значений детерминированных параметров — $(B_0^r, B_1^r, \ldots, B_k^r, \ldots, B_K^r)$, $r = 1, 2$, т. е. каждого из двух детерминированных значений во всех сечениях.

В качестве критериев оптимальности аппроксимации может быть принято минимальное абсолютное значение суммарной величины отклонений расчетных и фактических значений выходной переменной в каждом из сечений соответствующих нечетких множеств, т. е. величины $\min\left| a_k^r(\bar{Y}_i) - \sum_{j=1}^{n} b^j a_k^r(\bar{X}_k^j) - B_k^r \right|$, $r = 1, 2$, $k = 0, 1, ..., K$. Здесь $r = 1, 2$ — индексы крайних точек соответствующих сечений.

Задача выбора коэффициентов уравнения регрессии может быть представлена в виде задачи многокритериальной оптимизации вида

$$a_{k,1}^r(F) =$$
$$= \min_{\substack{b^1,\ldots,b^j,\ldots,b^n,\\ B_0^1,B_0^2,\ldots,B_K^1,B_K^2}} \sum_{i=1}^{N}\left[a_k^r(\bar{Y}_i) - \sum_{j=1}^{n} b^j a_k^r(\bar{X}_i^j) - B_k^r\right], \quad (9)$$
$$r = 1, 2,\ k = 0, 1, \ldots, K;$$
$$a_{k,2}^r(F) =$$
$$= \min_{\substack{b^1,\ldots,b^j,\ldots,b^n,\\ B_0^1,B_0^2,\ldots,B_K^1,B_K^2}} \sum_{i=1}^{N}\left[\sum_{j=1}^{n} b^j a_k^r(\bar{X}_i^j) + B_k^r - a_k^r(\bar{Y}_i)\right], \quad (10)$$
$$r = 1, 2,\ k = 0, 1, \ldots, K.$$

Аддитивная свертка критериев [17, 18] может быть представлена в виде

$$\bar{\Phi} = \min_{\substack{b^1,b^2,\ldots,b^k,\ldots,b^K,\\ B_0^1,B_0^2,\ldots,B_K^1,B_K^2}} \sum_{r=1}^{2}\left\{\delta_k^r \sum_{k=0}^{K}\sum_{i=1}^{N}\left[a_k^r(\bar{Y}_i) - \right.\right.$$
$$\left. - \sum_{j=1}^{n} b^j a_k^r(\bar{X}_i^j) - B_k^r\right] + \quad (11)$$
$$\left. + \mu_k^r \sum_{k=0}^{K}\sum_{i=1}^{n}\left[\sum_{j=1}^{n} b^j a_r^k(\bar{X}_i^j) + B_k^r - a_r^k(\bar{Y}_i)\right]\right\}.$$

Выражение (11) позволяет свести сформулированную многокритериальную задачу к минимизации одного критерия, линейного относительно вектора переменных задачи $(b^0, b^1, \ldots, b^j, \ldots, b^n, B_0^1, B_0^2, \ldots, B_K^1, B_K^2)$ в условиях выполнения ограничений на знаки соответствующих значений коэффициентов.

В выражении (11) приняты следующие обозначения: $0 \leqslant \delta_k^1 \leqslant 1$, $0 \leqslant \delta_k^2 \leqslant 1$; $0 \leqslant \mu_k^1 \leqslant 1$, $0 \leqslant \mu_k^2 \leqslant 1$ — весовые коэффициенты, которые удовлетворяют соотношению

$$\sum_{k=0}^{K}(\delta_k^1 + \delta_k^2 + \mu_k^1 + \mu_k^2) = 1; \quad (12)$$

$a_k^1(\bar{Y}_i),\ a_k^2(\bar{Y}_i);\ a_k^1(\bar{X}_k^1),\ a_k^2(\bar{X}_k^2),\ k = 0, 1, \ldots, K$ — соответственно минимальные и максимальные значения координаты абсцисс fuzzy-множеств выходной переменной и j-й входной переменной в k-м сечении i-го комплекта информации (исходных данных);

$B_k^1,\ B_k^2$ — соответственно значения левой и правой координат абсциссы fuzzy-множества свободного члена уравнения регрессии в k-м сечении.

Из свойств LR-представления функции принадлежности нечеткого множества уравнения регрессии должны быть предусмотрены следующие ограничения на коэффициенты этого уравнения:

$$\sum_{j=1}^{n} b^j a_{k-1}^1(\bar{X}^j) + B_{k-1}^1 \geqslant \sum_{j=1}^{n} b^j a_k^1(\bar{X}^j) + B_k^1; \quad (13)$$

$$\sum_{j=1}^{n} b^j a_{k-1}^2(\bar{X}^j) + B_{k-1}^2 \leqslant \sum_{j=1}^{n} b^j a_k^2(\bar{X}^j) + B_k^2. \quad (14)$$

Обозначим коэффициенты уравнения регрессии:

$$Q^j = \sum_{i=1}^{N}\sum_{k=0}^{K}[\delta_k^2 a_k^2(\bar{X}_i^j) - \mu_k^1 a_k^1(\bar{X}_i^j)];$$
$$D^j = \sum_{i=1}^{N}\sum_{k=0}^{K}[\mu_k^1 a_k^1(\bar{Y}_i) - \delta_k^2 a_k^2(\bar{Y}_i)],\ j = 1, \ldots, n. \quad (15)$$

Если все входные переменные имеют положительное влияние на выходную переменную задачи, т. е. с увеличением значения $\bar{X}^j$ значение $\bar{Y}$ должно увеличиваться, то задача аппроксимации сводится к решению только одной задачи линейного программирования — минимизации линейного критерия вида

$$\min_{\substack{b^1,\ldots,b^j,\ldots,b^n,\\ B_0^1,B_0^2,\ldots,B_K^1,B_K^2}} \sum_{j=1}^{n} Q^j b^j + \sum_{k=0}^{K}(\mu_k^1 B_k^1 - \delta_k^2 B_k^2); \quad (16)$$

$$b^j \geqslant 0,\ j = 1, \ldots, n;\ B_k^r \geqslant 0, k = 0, 1, \ldots, K,\ r = 1, 2. \quad (17)$$

В ряде случаев экспертами определены качественные влияния (положительные или отрицательные) входных параметров на выходную переменную, которые должны быть отражены также в уравнении регрессии. Пусть $\tilde{J}_1, \tilde{J}_2, \tilde{J}_3$ — соответственно подмножества входных переменных, для которых должно быть отражено положительное, отрицательное влияние входного фактора на выходную переменную и факторов, для которых характер влияния не установлен:

$$\tilde{J}_1 \cup \tilde{J}_2 \cup \tilde{J}_3 = \tilde{J} = \{j = 1, \ldots, n\},$$
$$J_1 \cap \tilde{J}_2 = \tilde{J}_1 \cap \tilde{J}_3 = \tilde{J}_2 \cap \tilde{J}_3 = \varnothing. \quad (18)$$

Пусть подмножество $\tilde{J}_3$ состоит из m_3 переменных, тогда оптимальное решение задачи может быть получено в результате решения 2^{m_3} задач вида

$$\tilde{\Phi}(\tilde{H}^3) = \min_{\substack{b^1,\ldots,b^n,\\ B_0^1,B_0^2,\ldots,B_K^1,B_K^2}} \sum_{j \in \tilde{J}^1 \cup \tilde{H}^3}\left\{Q^j b^j + \right.$$
$$\left. + \sum_{\tilde{J}^2 \cup (\tilde{J}^3 / \tilde{H}^3)} D^j b^j + \sum_{k=0}^{K}(\mu_k^1 B_k^1 - \delta_k^2 B_k^2)\right\} \quad (19)$$

в условиях ограничений (13), (14), 17).

Здесь $\tilde{H}^3$ — различные подмножества переменных из подмножества $\tilde{J}^3$, число которых равно 2^{m_3}. Среди этих всех полученных решений выбирается решение с наименьшим значением критерия оптимальности. Рассчитанные значения детерминированных коэффициентов этой математической модели b^j, $j = 1, ..., n$, и крайние точки соответствующих сечений нечеткого множества свободного члена B_k^r, $k = 0, 1, ..., K$, $r = 1, 2$, принимаются в виде нечеткой регрессионной модели, построенной на основе данного комплекта исходных данных.

4. Оценка адекватности Fuzzy-регрессионой модели

Результат расчета выходных показателей на основе нечеткой регрессионной модели — это fuzzy-множество $F(\bar{Y}_i)$, функция принадлежности которого $\mu_{\bar{Y}_i}[F(\bar{Y}_i)]$ представлена также многоугольником заданного вида (см. рисунок). В качестве детерминированного аналога прогнозируемой величины в каждом комплекте информации может использоваться координата абсциссы центра тяжести расчетного нечеткого множества — $G[F(\bar{Y}_i)]$.

В качестве оценки адекватности построенной Fuzzy-регрессионой модели могут быть приняты среднеквадратические значения суммы квадратов расчетного и фактического отклонений:

- координата абсцисс центра тяжести нечеткого множества, полученного в результате расчета по fuzzy-регрессионной модели $\{G[\bar{Y}_i] - G[F(\bar{Y}_i)]\}^2$, где $G[\bar{Y}_i]$ — фактическое значение абсциссы центра тяжести нечеткого множества значения выходной переменной, которое для $i = 1, ..., N$ вычисляется по формулам

$$G(\bar{Y}_i) = \frac{\int_{-\infty}^{\infty} \bar{Y}_i \mu_{\bar{Y}_i}(\bar{Y}_i) d\bar{Y}_i}{\int_{-\infty}^{\infty} \mu_{\bar{Y}_i}(\bar{Y}_i) d(\bar{Y}_i)};$$

- некоторая средневзвешенная величина абсолютных значений разности координат функции принадлежности фактического и расчетного значений функций принадлежности нечеткого множества $[Q[\bar{Y}_i] - Q[F(\bar{Y}_i)]]$. Здесь

$$Q[\bar{Y}_i] = \sum_{r=1}^{2}\sum_{k=0}^{K} \rho_k^r a_k^r [\bar{Y}_i];$$

$$Q[F(\bar{Y})_i] = \sum_{r=1}^{2}\sum_{k=0}^{K} \rho_k^r a_k^r F[(\bar{Y}_i)],$$

где $a_k^r[\bar{Y}_i]$ и $a_k^r[F(\bar{Y}_i)]$ — соответственно левые и правые крайние точки координат оси абсцисс функции принадлежности нечеткого множества, построенного по уравнениям регрессионной модели для i-го комплекта информации; $Q[\bar{Y}_i]$, $Q[F(\bar{Y})_i]$ — соответственно фактическое и расчетное значения функций принадлежности нечеткого множества в i-м комплекте информации; ρ_k^r, $r = 1, 2$, — весовые коэффициенты, значения которых в частном случае могут быть приняты такими же, как при расчете параметров регрессионной модели;

- абсолютное значение суммы отклонений расчетного и фактического значений функций принадлежности одних и тех же контрольных точек выходной переменной $E[\mu_{\bar{Y}_i}(\bar{Y}_i)] - E[\mu_{F(\bar{Y}_i)} F(\bar{Y}_i)]$, где

$$E[\mu_{\bar{Y}_i}(\bar{Y}_i)] = \sum_{r=1}^{2}\sum_{k=0}^{K} \delta_k^r \mu_{a_k^r(\bar{Y}_i)} [a_k^r(\bar{Y}_i)];$$

$$E[\mu_{\bar{Y}_i}\{F(\bar{Y}_i)\}] = \sum_{r=1}^{2}\sum_{k=0}^{K} \delta_k^r \mu_{a_k^r[F(\bar{Y}_i)]} \{[a_k^r[F(\bar{Y}_i)]\}.$$

Здесь $E[\mu_{\bar{Y}_i}(\bar{Y}_i)]$, $E[\mu_{\bar{Y}_i}\{F(\bar{Y}_i)\}]$ — соответственно значения фактической и расчетной сумм значений функций принадлежности одних и тех же контрольных точек выходной переменной.

В качестве оценки качества прогнозирования на основе нечеткой регрессионной модели может рассматриваться следующий показатель:

$$D = \frac{1}{N-1}\sum_{i=1}^{N} \{W(\bar{Y}_i) - W[F(\bar{Y}_i)]\}^2,$$

где в качестве значения $W(\bar{Y}_i)$ могут использоваться определенные выше показатели $G(\bar{Y}_i)$, $Q(\bar{Y}_i)$ или $E(\bar{Y}_i)$, а в качестве значения $W[F(\bar{Y}_i)]$ — показатели $G[F(\bar{Y}_i)]$, $Q[F(\bar{Y}_i)]$ или $E[F(\bar{Y}_i)]$.

С достаточной для практических приложений точностью в большинстве случаев могут использоваться построенные Fuzzy-регрессионные модели для прогнозирования значения выходной переменной, если справедливы следующие показатели их адекватности:

$$D \leqslant \omega \frac{1}{N-1}\{W(\bar{Y}_i^r) - M[W(\bar{Y}^r)]\}^2,$$

где $M[W(\bar{Y}^r)] = \frac{1}{N}\sum_{i=1}^{N} W(\bar{Y}_i^r)$, а значение весового коэффициента выбирается из условий $\omega \leqslant 0{,}1$.

Заключение

В условиях, когда в статистической выборке некоторые или все входные и выходная переменная представлены нечеткими множествами, а также из экономических соображений, технических характеристик или технологических условий установлены качественные влияния отдельных входных факторов, использование предлагаемых в работе fuzzy-регрессионных моделей является эффективной альтернативой получения количественных зависимостей установленных экспертами качественных закономерностей изучаемых явлений.

В работе предложены методы решения задачи в условиях, когда входные и выходная переменная и свободный член уравнения регрессии представлены fuzzy-множествами с LP-представлением функции принадлежности самого общего вида, а коэффициенты регрессии — действительные числа. В качестве критериев аппроксимации использованы сумма квадратов средневзвешенных координат минимальных и максимальных значений абсцисс сечений функций принадлежности нечетких множеств выходной переменной и их оценок по fuzzy-регрессионной модели. Построен детерминированный эквивалент и приведена вычислительная схема алгоритма решения задачи.

Ограничения, связанные с учетом необходимости учета в уравнении регрессионной модели положительного или отрицательного влияния на выходную переменную отдельных входных факторов, а также свойства арифметических операторов fuzzy-множеств, обусловили необходимость решения 2^{m_3} задач линейного программирования с последующим выбором среди них решения с наилучшим значением критерия оптимальности.

Полученные в работе результаты расширят область приложения fuzzy-регрессионных моделей в экономике, технических системах, социологии, маркетинге и других приложениях.

Список литературы

1. **Дрейпер Н., Смит Г.** Прикладной регрессионный анализ. Множественная регрессия. М.: Диалектика, 2007. 912 с.
2. **Орлова И. В., Половников В. А.** Экономико-математические методы и модели. М.: Вузовский учебник: ИНФРА-М, 2010. 366 с.
3. **Орлов А. И.** Нечисловая статистика. М.: М3-Пресс, 2004. 513 с.
4. **Зак Ю. А.** Принятие эффективных решений в экономике и менеджменте в условиях наличия нечисловой информации и размытых данных. М.: Экономика, 2018. 239 с.
5. **Зак Ю. А.** Методы обработки нечисловой информации в маркетинговых исследованиях // Маркетинг и маркетинговые исследования, Grebennikov. 2013. № 1. С. 20—33.
6. **Chang Yun-Hsi O.** Fuzzy regression methods — a comparative assessment // Fuzzy Sets und Systems. 2001. Vol. 119 (2). P. 187—203.
7. **Chang Yun-Hsi O.** Hybrid fuzzy least squares regression analysis and its reliability measures // Fuzzy Sets und Systems. 2001. Vol. 119 (2). P. 225—246.
8. **Штовба С. Д.** Нечеткая идентификация на основе регрессионных моделей параметрической функцией принадлежности // Проблемы управления и информатики. 2006. № 6. С. 1—8.
9. **Diamond P.** Fuzzy least squares // Information Sciences. 1988. Vol. 46. P. 141—157.
10. **Diamond P.** Least Squares Fitting of Several Fuzzy Variables // Proceedings of Secon IFSA Congress. Tokio, 1987. P. 20—25.
11. **Tanaka H., Uejima S., Asai K.** Linear regression analysis with fuzzy model // IEEE Trans. Systems Man Cybernet. 1982. Vol. 12, N. 6. P. 903—907.
12. **Tanaka H., Warada J.** Possiblistic linear system and their application tot he linear regression model // Fuzzy Sets und Systems. 1988. Vol. 27. P. 275—289.
13. **Celmins A.** Least Squares Model Fitting to Fuzzy Vektor Data // Fuzzy Sets und Systems. 1987. Vol. 22. P. 260—269.
14. **Rousseeuw P.** Applyng robust regression to insurance // Insurance: Mathematics and Economics. 1984. Vol. 3, N. 1. P. 67—72.
15. **Yang M.-S., Lee H. H.** Fuzzy Least Squares Alqorithmus for interactive Fuzzy Linear Regressions Models // Fuzzy Sets und Systems. 2003. Vol. 135, N. 2. P. 305—316.
16. **Aliev R., Fazlollahi B., Vahidov R.** Genetic algorithms-based fuzzy regression analysis // Soft Computing. 2002. N. 6. P. 470—475.
17. **Куни Р. Л., Райфа Х.** Принятие решений при многих критериях: предпочтения и замещения. М.: Радио и связь, 1981. 560 с.
18. **Зак Ю. А.** Прикладные задачи многокритериальной оптимизации. М.: Экономика, 2014. 455 с.
19. **Зак Ю. А.** Fuzzy-регрессионные модели в условиях наличия в статистической выборке нечисловой информации // Системнi дослiдження та iнфомацiйнi технологii. 2017. № 1. С. 88—96.

Yu. A. Zack, D. Sc., e-mail: yuriy_zack@hotmail.com

Fuzzy Linear Regression Analysis, Taking into Account the Influence of Input Factors

A solution is presented for fuzzy regression analysis problems in conditions where the input and output variables are represented by normalized fuzzy sets with an LR representation of the most general form membership function, and the regression coefficients are negative or positive real numbers. The free term of the regression equation is a fuzzy set of the most general form. There are restrictions on the degree of influence of some input factors established by experts. The approximation criterion is the minimum absolute value of the average unweighted sum of the absolute values of the coordinates of the minimum and maximum abscissa values of the cross sections for the membership functions of fuzzy sets of the output variable and its evaluation by the fuzzy regression model. Regression coefficients are calculated as a result of solving a certain subset of linear programming problems with the subsequent choice among them of the solution with the best value of the optimality criterion.

Keywords: *fuzzy regression models, fuzzy-sets, LR-representation of the membership function, cross sections of membership functions, linear programming*

DOI: 10.17587/it.26.137-144

References

1. **Drajper N., Smit G.** Applied Regression Analysis, Moscow, Dialektika, 2007, 912 p. (in Russian).
2. **Orlova I. V., Polovnikov V. A.** Economic and mathematical methods and models, Moscow, Busovskij uchebnik, INFRA-M, 2010, 366 p. (in Russian).
3. **Orlov A. I.** Non-numeric statistics, Moscow, M3-Press, 2004, 513 p. (in Russian).
4. **Zack Yu. A.** Making effective decisions in economics and management in the presence of non-numerical information and blurry data, Moscow, Ekonomika, 2018, 239 p. (in Russian).
5. **Zack Yu. A.** Non-numerical information processing methods in marketing research, *Marketing i marketingovije issledovanija, Grebennikov*, 2013, no. 1, pp. 20—33 (in Russian).
6. **Chang Yun-Hsi O.** Fuzzy regression methods — a comparative assessment, *Fuzzy Sets und Systems*, 2001, vol. 119 (2), pp. 187—203.
7. **Chang Yun-Hsi O.** Hybrid fuzzy least squares regression analysis and its reliability measures, *Fuzzy Sets und Systems*, 2001, vol. 119 (2), pp. 225—246.
8. **Shtovba S. D.** Fuzzy identification based on regression models by a parametric membership function, *Problemi Upravlenija i Informatiki*, 2006, no. 6, pp. 1—8 (in Russian).
9. **Diamond P.** Fuzzy least squares, *Information Sciences*, 1988, vol. 46, pp. 141—157.
10. **Diamond P.** Least Squares Fitting of Several Fuzzy Variables, Proceedings of Secon IFSA Congress, Tokio, 1987, pp. 20—25.
11. **Tanaka H., Uejima S., Asai K.** Linear regression analysis with fuzzy model, *IEEE Trans. Systems Man Cybernet*, 1982, vol. 12, no. 6, pp. 903—907.
12. **Tanaka H., Warada J.** Possiblistic linear system and their application tot he linear regression model, *Fuzzy Sets und Systems*, 1988, vol. 27, pp. 275—289.
13. **Celmins A.** Least Squares Model Fitting to Fuzzy Vektor Data, *Fuzzy Sets und Systems*, 1987, vol. 22, pp. 260—269.
14. **Rousseeuw P.** Applyng robust regression to insurance, *Insurance: Mathematics and Economics,* 1984, vol. 3, no. 1, pp. 67—72.
15. **Yang M.-S., Lee H. H.** Fuzzy Least Squares Alqorithmus for interactive Fuzzy Linear Regressions Models, *Fuzzy Sets und Systems*, 2003, vol. 135, no. 2, pp. 305—316.
16. **Aliev R., Fazlollahi B., Vahidov R.** Genetic algorithms-based fuzzy regression analysis, *Soft Computing*, 2002, no. 6, pp. 470—475.
17. **Kuni R. L., Rajfa Ch.** Decision making under many criteria: preferences and substitutions, Moscow, Radio i Svyaz, 1981, 560 p. (in Russian).
18. **Zack Yu. A.** Applied problems of multicriteria optimization, Moscow, Ekonomika, 2014, 455 p. (in Russian).
19. **Zack Yu. A.** Fuzzy-regression models in the presence of non-numerical information in a statistical sample, *Systemni doslidzhennya ta infomatsiyni tekhnolohii*, 2017, no. 1, pp. 88—96 (in Russian).