УДК 004

Исследование методов оптимизации для задач нечеткой регрессии

Нежильской Михаил Вячеславович
Волжский политехнический институт (филиал) «Волгоградский государственный технический университет»
Студент

Фадеева Марина Викторовна
Волжский политехнический институт (филиал) «Волгоградский государственный технический университет»
Старший преподаватель

Рыбанов Александр Александрович
Волжский политехнический институт (филиал) «Волгоградский государственный технический университет»
к.т.н., доцент, зав. кафедрой информатика и технология программирования

Аннотация
Рассмотрены алгоритмы решения задач нечеткого регрессионного анализа в условиях, когда входные и выходные переменные представлены нечеткими множествами, определенными с точностью до неизвестных параметров, а коэффициенты регрессии — действительные числа. Описаны алгоритмы преобразования переменных, представленных термами лингвистической переменной или параметрами числовых шкал, в нечеткие множества и использования этих данных в задачах нечеткого регрессионного анализа. Рассмотрены возможности использования различных моделей и методов принятия решений в условиях неопределенности, которая имеет разные интерпретации и обусловлена не только влиянием внешней среды, но и характеристиками исходной информации. Полученные результаты позволят решать многие прикладные проблемы в экономике, логистике, социологии, маркетинге, техники, проектирования и других областях.
Ключевые слова: нечеткий регрессионного анализ, нечеткие множества, нечеткая информация, многокритериальность, оптимальное значение

Research of methods optimization for problems of fuzzy regression

Nezhilskoy Mikhail Vyacheslavovich
Volzhskiy Polytechnical Institute, branch “Volgograd State Technical University”
student

Fadeeva Marina Viktorovna
Volzhskiy Polytechnical Institute, branch “Volgograd State Technical University”
scientific director

Rybanov Aleksandr Aleksandrovich
Volzhskiy Polytechnical Institute, branch “Volgograd State Technical University”
Candidate of Engineering Sciences, associate professor, Head of the Department of Informatics and programming technology

Abstract
Algorithms of a solution of problems of fuzzy regression analysis in conditions when input and output variables are presented by the fuzzy sets defined to within unknown parameters, and regression coefficients — real numbers are considered. Algorithms of conversion of the variables presented by terms of a linguistic variable or parameters of numerical scales in fuzzy sets and uses of these data in problems of fuzzy regression analysis are described. The possibilities of use of different models and methods of decision-making in the conditions of uncertainty which has different interpretations are considered and it is caused not only by influence of the external environment, but also characteristics of an original information. The received results will allow to solve many applied problems in economy, logistics, sociology, marketing, the equipment, design and other areas.
Keywords: fuzzy regression analysis, fuzzy sets, fuzzy information, optimum value

Цель работы - исследование методов нечёткого регрессионного анализа и алгоритмов оптимизации задач нечёткой регрессии.

В наше время большую значимость в организационной деятельности российских и зарубежных кампаний занимает количественный анализ результатов эффективности производства. Это влечет за собой спрос на разработку удобных программных продуктов, новых способов анализа финансовой и общеэкономической информации [1].

На сегодняшний день одним из наиболее перспективных направлений научных исследований в области прогнозирования, анализа и моделирования социальных, экономических и маркетинговых процессов и явлений является нечеткая логика (fuzzy logic) [4]. С помощью нечетких множественных моделей можно проводить исследования на очень высоком уровне вне зависимости от точности и полноты имеющейся информации, способствуют принятию более обоснованных решений возникающих задач. Внедрение в работу предприятий научных разработок, основанных на теории нечетких множеств, позволяют улучшить качество их работы, сделать более "мобильными" при анализе быстро меняющейся информации.

Одной из наиболее распространенных и изученных форм обработки и исследования информации является регрессионный анализ. В ситуациях, когда многие входные факторы модели могут быть представлены лишь булевыми, лингвистическими или нечеткими данными, либо некоторыми градациями числовых шкал, в качестве альтернативных подходов могут использоваться методы нечеткого регрессионного анализа [2]. Результатом расчета на основе математических моделей нечеткого регрессионного анализа является некоторое нечеткое множество с функцией принадлежности непрерывного вида, которое определяет диапазон возможных значений выходной переменной и оценку получения этого значения в пределах данного диапазона. Это дало возможность применять методы нечеткого регрессионного анализа в области теории экспертного оценивания, экономики и маркетинге.

Построение нечеткой регрессионной модели состоит в нахождении оптимальных в каком-то смысле коэффициентов с учетом нечеткой информации об объекте и субъективных представлений исследователя [3].

Базовые предположения нечеткой регрессии заключаются в том, что остатки, полученные как разность между наблюдениями и их оценками, продуцируются не случайными ошибками измерения, а неопределенностями (типа нечеткость) при вычислении параметров модели.

Выделяют два основных подхода к построению моделей нечеткой линейной регрессии [5]:

1. нечеткая регрессия, основанная на критерии минимизации нечеткости;

2. подход, комбинированный с методом наименьших квадратов.

Рассмотрим задачу нечёткого регрессионного анализа, в которой имеется матрица наблюдений, каждая из строк которой – комплект входной и выходной информации. Матрица представлена нечеткими множествами. Таблица исходных данных состоит из N комплектов экспериментальных данных. Необходимо найти нечеткую линейную регрессионную модель вида

$$\bar{Y} = a_1 \otimes \bar{X}_1 + a_2 \otimes \bar{X}_2 + \ldots + a_j \otimes \bar{X}_{\mathrm{j}} + \cdots + a_n \otimes \bar{X}_{\mathrm{n}} + \bar{a}_0,$$

где $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_{\mathrm{j}}, \ldots, \bar{X}_{\mathrm{n}}$; $\bar{a}_0$ — некоторые нечеткие множества с заданными функциями принадлежности (в частном случае некоторые из них — действительные числа), а коэффициенты модели $a_1, a_2, \ldots, a_{\mathrm{j}}, \ldots, a_{\mathrm{n}}$ — некоторые действительные числа.

Рассмотрим математическую модель этой задачи в случае, когда нечёткие множества входных и выходных переменных представлены треугольными функциями принадлежности. Центральные точки функций принадлежности $\mu(\bar{X}_{ij})$ и $\mu(\bar{Y}_i)$ соответственно входных $\bar{X}_{ij}$ и выходных переменных $\bar{Y}_i$ этих нечётких множеств обозначим соответственно m_{ij} и $\bar{m}_i$, левые крайние точки — $\lambda_{ij}, \bar{\lambda}_i$, а правые крайние точки — γ_{ij} и $\bar{\gamma}_i$, где $j = 1, .., n, i = 1, .., N$.

Критерий качества аппроксимации — минимум средневзвешенной суммы квадратов отклонений расчетных параметров выходной переменной по нечёткой регрессионной модели от их фактических значений:

$$\Phi_1 = \eta_1 \sum_{i=1}^{N} \left[\overline{m}_i - \left(\sum_{j=1}^{N} a_j m_{ij} + a_0 \right) \right]^2 + \eta_2 \sum_{i=1}^{N} \left[\bar{\lambda}_i - \left(\sum_{j=1}^{N} a_j \lambda_{ij} + a_0 \right) \right]^2 + \eta_3 \sum_{i=1}^{N} \left[\bar{\gamma}_i - \left(\sum_{j=1}^{N} a_j \gamma_{ij} + a_0 \right) \right]^2 \to min$$

η_1, η_2, η_3, где $0 < \eta_r \leq 1$, $r = 1, 2, 3$ — весовые коэффициенты, определяющие важность значения каждого из параметров, причем $\eta_1 + \eta_2 + \eta_3 = 1$.

Координата абсцисс центра тяжести нечеткого множества для функций принадлежности треугольного типа имеет вид:

$$G(\overline{X}_{ij}) = \frac{1}{3}(\lambda_{ij} + m_{ij} + \gamma_{ij});\ G(\overline{Y}_i) = \frac{1}{3}(\bar{\lambda}_i + \overline{m}_i + \bar{\gamma}_i), \qquad j = 1,..,n; i = 1,..,N,$$

$$\Phi_2 = \sum_{i=1}^{N} \left[G(\overline{Y}_i) - \left(\sum_{j=1}^{N} a_j G(\overline{X}_{ij}) + a_0 \right) \right]^2 = \sum_{i=1}^{N} \left[(\bar{\lambda}_i + \overline{m}_i + \bar{\gamma}_i) - \left(\sum_{j=1}^{N} a_j (\lambda_{ij} + m_{ij} + \gamma_{ij}) + a_0 \right) \right]^2 \to min$$

$$\Phi_3 = \sum_{i=1}^{N} \left[Y_i - \left(\sum_{j=1}^{N} a_j z_{ij} + a_0 \right) \right]^2 \to min,$$

где для треугольных функций принадлежности значения z_{ij} $j = 1,..,n; i = 1,..,N$, вычисляются соответственно по формулам:

$$z_{ij} = \sum_{p=1}^{P} 0{,}5\omega_p \{ [m_{ij} - (m_{ij} - \lambda_{ij})(1-\beta)] + [m_{ij} + (\gamma_{ij} - m_{ij})(1-\beta)] \}$$

где $0 \leq \omega_p \leq 1, p = 1, \dots, P$ — весовые коэффициенты, удовлетворяющие соотношению $\sum_{p=1}^{P} \omega_p = 1$; $0 \leq \beta_p \leq 1{,}0, p = 1, \dots, P$ — значения функции принадлежности в различных сечениях.

Необходимыми и достаточными условиями достижения локального минимума значения критерия Φ_1 является система линейных уравнений вида

$$\frac{d\Phi_p}{da_0} = 0, \frac{d\Phi_p}{da_j} = 0, p = 1, 2, 3, 4; j = 1, \dots, n.$$

Вычислив частные производные по каждому из оптимизируемых параметров и выполнив ряд алгебраических преобразований, получим следующую систему линейных алгебраических уравнений для определения значений коэффициентов $a_1,\ a_2, \dots, a_j, \dots, a_n$ —

$$\sum_{j=1}^{\mathrm{n}} \beta_{kj} a_j + a_0 = \mathrm{B_k}, \mathrm{k} = 1, \dots, \mathrm{n}.$$

Перейдем к нормированным показателям. Вычислим

$$M(z_j) = \frac{1}{N}\sum_{i=1}^{N} z_{ij}, \sigma^2(z_j) = \frac{1}{N-1}\sum_{i=1}^{N}\left[\sum_{i=1}^{N} z_{ij} - M(z_j)\right]^2, j = 1, \dots, n;$$

$$M(Y) = \frac{1}{N}\sum_{i=1}^{N} Y_i, \sigma^2(Y) = \frac{1}{N-1}\sum_{i=1}^{N}\left[\sum_{i=1}^{N} Y_i - M(Y_i)\right]^2;$$

$$\rho_{kj} = \frac{\frac{1}{N}\sum_{i=1}^{N}[z_k - M(z_k)][z_j - M(z_j)]}{\sigma(z_k)\sigma(z_j)}, j, k = 1, \dots, n, j \neq k;$$

$$\rho_{kk} = 1;\ \rho(y, z_k) = \frac{\frac{1}{N}\sum_{i=1}^{N}[z_{ki} - M(z_k)][Y_i - M(Y)]}{\sigma(z_k)\sigma(Y)}, k = 1, \dots, n,$$

и перейдем к решению следующей системы n линейных алгебраических уравнений относительно переменных $\delta_j,\ j = 1, \dots, n$:

$$\sum_{j=1}^{n} \rho_{kj}\delta_j = \rho(y, z_k), k = 1, \dots, n.$$

Вычислив значения коэффициентов $\delta_j,\ j = 1, \dots, n$, определим значения коэффициентов $\bar{a}_j,\ j = 1, \dots, n$, в нормальном масштабе измерения по формулам $\bar{a}_j = \delta_j \cdot \frac{\sigma(Y)}{\sigma(z_j)}, j = 1, \dots, n.$

Оптимальные значения факторов нечеткой регрессионной модели определяются с помощью аппарата многокритериальной оптимизации[6].

Рассмотрим задачу многокритериальной оптимизации следующего вида:

$$\begin{cases} y^1 = f_1(x) \to max, \\ \dots \\ y^n = f_n(x) \to max, \\ x \in X, \end{cases}$$

где $x = (x_{i_1}, \dots, x_{i_n}) \in R^n, X \in R^n$ – множество допустимых решений (факторов модели); $\forall i = \overline{1,n}\ y = (f_1(x), \dots, f_n(x))$; $(f_i: R^n \to R)$ – целевые функции (критерии).

Пусть все функции $y^i = f_i(x) (i = \overline{1,N})$ являются непрерывно-дифференцируемыми в X, тогда может быть определен градиент в любой точке $x \in X$ для каждой целевой функции

$$\nabla y^i = \left(\frac{\partial y^i}{\partial x_{i_1}}, \dots, \frac{\partial y^i}{\partial x_{i_n}}\right)^T -$$

вектор, указывающий направление, в котором увеличивается значение целевой функции.

Рассмотрим частный случай, в котором все целевые функции линейны, т.е.

$$\forall i = \overline{1, N} \left(y^i = \sum_{p=1}^{n_j} \beta_i^p \cdot x_{i_p} \right),$$

где $x = \left(x_{i_1}, \dots, x_{i_n}\right)^T \in X \in R^n, \beta^p = (\beta^1, \dots, \beta^{n_j})^T$ – вектор коэффициентов целевой функции.

В линейном случае градиент полностью определяется коэффициентами целевой функции $\nabla y^i = \beta_j$ и представляет собой константу.

Коэффициент взаимодействия целевых функций определяется по формуле

$$k_{ij} = cos\varphi = \frac{\sum_{p=1}^{n_j} \beta_i^p \cdot \beta_j^p}{\sqrt{\sum_{l=1}^{n_j} \left(\beta_i^l\right)^2} \cdot \sqrt{\sum_{l=1}^{n_j} \left(\beta_j^l\right)^2}}.$$

Для определения типа взаимодействия разобьем отрезок $[0, \pi]$ на три промежутка: $[0, \pi] = \left[0, \frac{\pi}{3}\right] \cup \left(\frac{\pi}{3}, \frac{2\pi}{3}\right) \cup \left[\frac{2\pi}{3}, \pi\right]$.

С учетом введенного разбиения на основе коэффициентов k_{ij} можно сформулировать следующие правила принятия решений:

1. чем ближе k_{ij} к 1, тем в большей степени целевые функции $f_i(x)$ и $f_j(x)$ являются кооперирующими, поэтому если $k_{ij} \in \left[\frac{1}{2}, 1\right]$, то цели кооперируют;
2. чем ближе k_{ij} к -1, тем в большей степени цели $f_i(x)$ и $f_j(x)$ конкурируют, поэтому если $k_{ij} \in \left[-1, -\frac{1}{2}\right]$, то цели конкурируют;
3. чем ближе k_{ij} к 0, тем в большей степени цели независимы, поэтому если $k_{ij} \in \left(-\frac{1}{2}, \frac{1}{2}\right)$, то цели независимы.

Вычислим коэффициенты взаимодействия для каждой пары целевых функций k_{ij}, а затем сформируем матрицу $K = \left\{k_{ij}\right\}_{N \times N}$ с элементами $\left|k_{ij}\right| \leq 1$. Знак конкретного коэффициента позволяет сделать вывод о типе взаимодействия, при этом также важна и количественная оценка такого взаимодействия.

Для оценки силы взаимодействия целей воспользуемся понятием нечеткого множества. Каждому из промежутков $\left[0, \frac{\pi}{3}\right] \cup \left(\frac{\pi}{3}, \frac{2\pi}{3}\right) \cup \left[\frac{2\pi}{3}, \pi\right]$ ставится в соответствие промежуток изменения косинуса соответствующего угла. Соответственно имеем $\left[\frac{1}{2}, 1\right], \left(-\frac{1}{2}, \frac{1}{2}\right), \left[-1, -\frac{1}{2}\right]$.

В настоящей статье рассмотрены подход построения нечетких регрессионных моделей, аппарат многокритериальной оптимизации и методы преобразования многих видов нечисловой информации в действительное число или нечеткое множество, а также использование этой информации в вычислительных схемах расчета детерминированных коэффициентов уравнения регрессии.

Опираясь на рассмотренный материал можно сделать следующие выводы:

1. в нечётких регрессионных моделях, в отличие от традиционных регрессионных зависимостей, рассчитанное значение выходной переменной представлено в виде некоторого диапазона возможных значений с оценкой веса каждого из этих значений в пределах этого диапазона, что позволит в ряде случаев более объективно оценить риск принимаемых решений на основе полученных результатов расчета;
2. в случае представления входных и выходных переменных нечисловой информацией (лингвистические переменные, булевы, данные числовых шкал, нечеткие множества и т.п.), использование нечётких регрессионных моделей является эффективной альтернативой получения количественных зависимостей, установленных экспертами качественных закономерностей изучаемых явлений;
3. использование аппарата многокритериальной оптимизации позволяет определить оптимальные значения факторов нечеткой регрессионной модели.

Полученные результаты позволят решать многие прикладные проблемы в экономике, логистике, социологии, маркетинге, техники, проектирования и других областях.

Библиографический список

1. Вишнякова Е.В., Иванова Е.В., Камалов С.М., Колодяжная Ю.А., Хамидуллина Л.Ф. Нечеткая линейная регрессия в задачах оценки// Научные записки молодых исследователей. 2015. № 5. С. 14–29.
2. Грицюк В.И. Нечеткий робастный регрессионный анализ для нечетких входных и выходных данных / В.И. Грицюк // Технологический аудит и резервы производства. 2015. № 6. С. 4–8.
3. Зак Ю.А. Математические модели прогнозирования затрат времени и стоимости перевозки грузов // Логистика сегодня, Grebennikov. 2015. №1. С. 162–172.
4. Зак Ю.А. Принятие решений в условиях размытых и нечетких данных // URSS, М., 2013. 352 с.
5. Згуровский М., Зайченко Ю. Модели и методы принятия решений в нечетких условиях. К.: Наук.думка, 2013. 275 с.
6. Мелькумова Е.М. (Аристова Е.М.) Многокритериальная оптимизация на основе меры зависимости целевых функций // Известия Тульского государственного университета. Естественные науки. 2011. №1. С. 177-187.