

Раздел 2 МОДЕЛИ, СИСТЕМЫ, СЕТИ В ТЕХНИКЕ

Section 2 MODELS, SYSTEMS, NETWORKS IN THE TECHNIQUE

УДК 519.862.6
doi:10.21685/2227-8486-2022-1-8

ПОСТРОЕНИЕ СВЕРТКИ КРИТЕРИЕВ АДЕКВАТНОСТИ РЕГРЕССИОННЫХ МОДЕЛЕЙ

С. И. Носков

Иркутский государственный университет путей сообщения, Иркутск, Россия
sergey.noskov.57@mail.ru

Аннотация. *Актуальность и цели.* В работе предложен способ построения линейной свертки критериев адекватности регрессионных уравнений на основе информации, сформированной в ходе проведения конкурса моделей. Рассмотрены случаи совместности и несовместности системы неравенств, порожденной этой информацией. В первом из них предлагается искать РС-решение системы, максимизирующее ее разрешающую способность, во втором – определять квазирешение, минимизирующее указанную несовместность. *Материалы и методы.* Для достижения цели привлечен аппарат решения задач линейного программирования со смешанными ограничениями. *Результаты.* В результате решения сформированных задач линейного программирования производится оценка коэффициентов линейной свертки частных критериев адекватности регрессионной модели, основанная на индивидуальных предпочтениях исследователя относительно сравнительной значимости этих критериев. *Выводы.* Построенная свертка критериев может быть многократно использована исследователем при построении моделей других объектов.

Ключевые слова: регрессионная модель, линейная свертка, критерии адекватности, РС-решение, квазирешение, задача линейного программирования

Для цитирования: Носков С. И. Построение свертки критериев адекватности регрессионных моделей // Модели, системы, сети в экономике, технике, природе и обществе. 2022. № 1. С. 73–81. doi:10.21685/2227-8486-2022-1-8

CONSTRUCTION OF CONVOLUTION CRITERIA FOR REGRESSION MODELS

S.I. Noskov

Irkutsk State Transport University, Irkutsk, Russia
sergey.noskov.57@mail.ru

Abstract. *Background.* The paper proposes a method for constructing a linear convolution of criteria for the adequacy of regression equations based on information generated during the competition of models. Cases of compatibility and inconsistency of the system of inequalities generated by this information are considered. In the first of them, it is proposed to search for a PC-solution of this system, which maximizes its resolution, and in the second, to determine a quasi-solution of the system that minimizes the indicated inconsistency. *Materials and methods.* To achieve the goal, an apparatus for solving linear programming problems with mixed constraints is involved. *Results.* As a result of solving the formed linear programming problems, the coefficients of the linear convolution of the partial criteria for the adequacy of the regression model are estimated based on the individual preferences of the researcher regarding the comparative significance of these criteria. *Conclusions.* The constructed convolution of criteria is based on the individual preferences of the researcher regarding the comparative significance of these criteria and can be repeatedly used by him when constructing models of other objects.

Keywords: regression model, linear convolution, adequacy criteria, PC-solution, quasi-solution, linear programming problem

For citation: Noskov S.I. Construction of convolution criteria for regression models. *Modeli, sistemy, seti v ekonomike, tekhnike, prirode i obshchestve = Models, systems, networks in economics, technology, nature and society.* 2022;(1):73–81. (In Russ.). doi:10.21685/2227-8486-2022-1-8

Введение

Общеизвестно, что регрессионные модели давно и успешно используются при анализе сложных объектов самого различного масштаба и характера функционирования. Так, в работе [1] описано применение скорректированной логистической регрессии для изучения связи между медицинским страхованием и использованием амбулаторных и стационарных медицинских услуг. В статье [2] применяется метод многомерных адаптивных регрессионных сплайнов для моделирования процесса подземной газификации угля – технологии, позволяющей добывать уголь без традиционного горнодобывающего оборудования. В [3] регрессионный анализ используется для оценки эффективности банков США с учетом их неоднородности. В весьма интересном исследовании [4] с помощью регрессионных методов доказывается правильность выбранных мер по выводу из кризиса экономики США в период Великой депрессии. В работе [5] производится выявление и оценка влияния детерминант развития занятости в неформальном секторе в регионах России с учетом влияния пространственных эффектов с использованием моделей регрессии и классических моделей панельных данных.

Одной из основных проблем, связанных с построением качественных регрессионных моделей, является комплексная оценка их адекватности с по-

мощью соответствующих критериев. В настоящей работе рассмотрен один из возможных подходов к ее решению.

Постановка задачи

Пусть перед исследователем при анализе сложного технического или социально-экономического объекта поставлена задача формализовать существующую связь между зависимой (выходной) переменной y и влияющими на нее независимыми (входными) переменными x_i , $i = \overline{1, m}$, т.е. построить математическую модель (уравнение, зависимость) регрессионного типа:

$$y_k = F(\alpha; x_{k1}, x_{k2}, \dots, x_{km}) + \varepsilon_k, k = \overline{1, n}, \quad (1)$$

где F – аппроксимирующая функция; α – подлежащий оцениванию вектор параметров; k – номер наблюдения; ε_k – ошибки аппроксимации; n – число наблюдений (длина выборки).

Пусть также при построении вполне адекватной объекту зависимости (1) исследователь проводит так называемый конкурс моделей [6], состоящий в формировании множества их альтернативных вариантов и последующем выделении лучшего из них, руководствуясь широким набором разработанных в рамках регрессионного анализа критериев адекватности. Причинами множественности указанных альтернатив могут, в частности, быть:

- варьирование вида аппроксимирующей функции F ;
- применение различных методов оценивания параметров модели (1), в том числе наименьших квадратов и модулей, антиробастного оценивания [6], максимальной согласованности [7] и многих других;
- расширение круга независимых переменных путем включения в него их различных преобразований.

Для оценки адекватности участвовавших в конкурсе вариантов модели могут быть привлечены критерии [6]:

- Дарбина – Уотсона;
- информативности набора независимых переменных;
- множественной детерминации;
- остаточной дисперсии;
- смещения;
- согласованности поведения расчетных и наблюдаемых значений зависимой переменной;
- средней относительной ошибки аппроксимации;
- средней относительной ошибки прогноза;
- Стьюдента;
- Фишера;
- ширины области определения уравнения (1).

Итак, пусть в результате проведения конкурса моделей исследователем сформировано множество альтернативных вариантов зависимости (1) $\Omega = \{M_1, M_2, \dots, M_r\}$, для оценки адекватности которых привлечено t критериев K_1, K_2, \dots, K_t . Будем полагать, что все критерии K_i , $i = \overline{1, t}$ приведены к однородному виду, т.е. значение каждого из них тем лучше, чем оно

выше. Приемы, позволяющие произвести соответствующие операции преобразования, подробно описаны в монографии [6].

Дальнейшая реализация конкурса может быть осуществлена на основе применения одной из вычислительных схем, описанных в работах [6, 8]. Каждая из них содержит этапы последовательного сужения множества Ω путем удаления из него вариантов, знаки параметров которых не соответствуют содержательному смыслу переменных, а значения критериев адекватности выходят за установленные пределы. Окончательный выбор лучшего варианта производится на основе использования часто применяемых в рамках теории решения многокритериальных задач методом уступок или идеальной точки.

Как отмечено в [6], современные системы статистических моделей, как правило, имеют большую размерность и содержат сотни или даже тысячи регрессионных уравнений. Исследователь при построении соответствующих зависимостей, реализуя для каждой из них конкурс моделей, пользуется обычно одной и той же (или аналогичной) системой приоритетов на множестве привлекаемых критериев, тем более, если строящиеся уравнения будут использоваться для решения одной проблемы, связанной, например, с краткосрочным прогнозированием. Это обстоятельство делает организацию конкурса моделей для однородных уравнений на основе, например, метода уступок, не вполне эффективной, поскольку она хоть и приводит к достаточно простому диалогу с ЭВМ, требует, тем не менее, определенного времени на отыскание лучшего варианта для каждого уравнения. Гораздо более эффективным является, по-видимому, проведение такого конкурса только для одного уравнения с тем, чтобы на основе полученной информации сформировать некое правило (индивидуальное для каждого разработчика), которое можно было бы легко применять для определения лучшего варианта других уравнений данной группы. Занесение такого сорта правил в базу знаний системы автоматизации моделирования в виде совокупности формализованных в соответствующих программных модулях предпочтений для каждого исследователя позволит значительно оперативнее разрабатывать крупные модельные комплексы, привнося в процесс моделирования дополнительный интеллектуальный элемент. При этом в зависимости от характера проблемы, для решения которой строится та или иная модель, а также группы, к которой относится описываемый показатель, каждый пользователь может применять свое индивидуальное правило, уже хранящееся в базе знаний. В случае, когда такое правило дает «сбой» на некотором уровне (т.е. приведет к построению неудовлетворительного варианта), его следует откорректировать, проведя конкурс моделей по «полной программе».

В правила, о которых идет речь, можно вкладывать различный смысл и формализовывать их по-разному. Представляется, что наиболее удобной и эффективной формой их представления является свертка критериев, участвующих в оценивании качества вариантов зависимости (1), а различаться правила между собой в этом случае должны значениями коэффициентов свертки.

Пусть в результате проведения конкурса моделей при построении уравнения (1) выделен лучший его вариант $M^* \in \Omega$. Без потери общности будем полагать, что $M^* = M_1$. Задача состоит в построении свертки линейных критериев K_i , $i = \overline{1, t}$ вида:

$$K(M) = \sum_{i=1}^t \beta_i K_i(M), \beta_i > 0, i = \overline{1, t}, \quad (2)$$

т.е. в определении коэффициентов $\beta_i, i = \overline{1, t}$.

Определение коэффициентов линейной свертки критериев адекватности регрессионных моделей

Рассмотрим две возможных ситуации.

1. Система линейных ограничений, порожденная сформированной в ходе проведения конкурса моделей информацией, является совместной.

В этом случае значение свертки критериев на варианте M_1 должно быть не меньше, чем на других вариантах:

$$K(M_1) \geq K(M_k), k = \overline{2, r},$$

или, с учетом (2),

$$\sum_{i=1}^t \beta_i K_i(M_1) \geq \sum_{i=1}^t \beta_i K_i(M_k), k = \overline{2, r}. \quad (3)$$

В соответствии с часто используемым в теории принятия решений приемом (см., например, [9]), потребуем, чтобы разрешающая способность системы неравенств (3) была максимальной, или, что то же, найдем ее РС-решение [10].

Для этого вначале обеспечим положительность коэффициентов свертки (2):

$$\beta_i \geq \gamma, i = \overline{1, t}, \quad (4)$$

где γ – наперед заданное малое положительное число.

Далее преобразуем систему неравенств (3) следующим образом:

$$\sum_{i=1}^t \beta_i (K_i(M_1) - K_i(M_k)) - u_k \geq 0, k = \overline{2, r}, \quad (5)$$

$$u_k \geq 0, k = \overline{2, r}. \quad (6)$$

Имеет смысл придать свертке (2) относительный характер, например, масштабировав ее значения на 100 % следующим образом:

$$\sum_{i=1}^t \beta_i \bar{K}_i = 100, \quad (7)$$

где

$$\bar{K}_i = \max_{k=1, r} K_i(M_k).$$

После этого максимизируем разрешающие поправки:

$$\sum_{k=2}^r u_k \rightarrow \max. \quad (8)$$

Решение задачи линейного программирования (ЛП) (4)–(8) позволяет определить искомые коэффициенты свертки (2).

2. Пусть теперь система ограничений (3) несовместна.

Одним из способов разрешения такой ситуации является отыскание такого вектора параметров $\beta = (\beta_1, \beta_2, \dots, \beta_t)$, который минимизирует степень этой несовместности, или, другими словами, является квазирешением такой системы [10].

Заменим ограничения (5), (6) на следующие:

$$\sum_{i=1}^t \beta_i (K_i(M_1) - K_i(M_k)) + v_k \geq 0, \quad k = \overline{2, r}, \quad (9)$$

$$v_k \geq 0, \quad k = \overline{2, r}, \quad (10)$$

а целевую функцию (8) на

$$\sum_{k=2}^r v_k \rightarrow \min. \quad (11)$$

Тогда коэффициенты линейной свертки критериев (2) определяются посредством решения задачи ЛП (4), (7), (9)–(11).

Численный пример

Пусть в результате проведения конкурса моделей в соответствии с описанной в [8] вычислительной схемой построено пять альтернативных вариантов уравнения (1). При этом были задействованы критерии множественной детерминации K_1 и выраженные в процентах согласованность поведения расчетных и наблюдаемых значений зависимой переменной (K_2) и ширина области определения уравнения (K_3). Заметим, что все три критерия «работают на максимум» и поэтому не требуют преобразования к однородному виду. Их значения для каждого варианта приведены в табл. 1.

Таблица 1

Значения критериев адекватности на вариантах модели

Номер варианта	K_1	K_2	K_3
1	0,86	50	80
2	0,90	51	79
3	0,85	70	70
4	0,82	35	90
5	0,85	55	75

Как нетрудно убедиться, система ограничений (4)–(7), сформированная на данной информации, оказывается несовместной. Следовательно, имеет место второй случай из приведенных выше. Рассмотрим его более подробно.

В соответствии с ограничениями (4) потребуем, чтобы значения параметров свертки (2) были снизу ограничены числом 0,001.

Ограничения (9) примут вид:

$$\begin{aligned} -0,04\beta_1 - \beta_2 + \beta_3 + v_2 &\geq 0; \\ 0,01\beta_1 - 20\beta_2 + 10\beta_3 + v_3 &\geq 0; \\ 0,04\beta_1 + 15\beta_2 - 10\beta_3 + v_4 &\geq 0; \\ 0,01\beta_1 - 5\beta_2 + 5\beta_3 + v_5 &\geq 0. \end{aligned}$$

Ограничение (7) примет форму

$$0,90\beta_1 + 70\beta_2 + 90\beta_3 = 100.$$

Решение задачи ЛП (4), (7), (9)–(11) позволяет сформировать линейную свертку критериев адекватности (2) в виде

$$K = 30,211K_1 + 0,302K_2 + 0,574K_3. \quad (12)$$

Пользуясь сверткой (12) и ограничением (7), определим оценки значимости $\delta_i, i = \overline{1,3}$ каждого из критериев:

$$\begin{aligned} \delta_1 &= 30,211 \cdot 0,90 = 27,19; \\ \delta_2 &= 0,302 \cdot 70 = 21,15; \\ \delta_3 &= 0,574 \cdot 90 = 51,66. \end{aligned}$$

Таким образом, наиболее значимым для исследователя является критерий K_3 , важность других двух критериев уступает ему примерно в два раза.

Заключение

В работе предложен способ построения линейной свертки критериев адекватности регрессионных уравнений на основе информации, сформированной в ходе проведения конкурса моделей. Рассмотрены случаи совместной и несовместной системы неравенств, порожденной этой информацией. В первом из них предлагается искать РС-решение этой системы, максимизирующее ее разрешающую способность, во-втором – определять квазирешение системы, минимизирующее указанную несовместность. Рассмотрен численный пример.

Список литературы

1. Tungu M., Amani P. J., Hurtig A.-K. [et al.]. Does health insurance contribute to improved utilization of health care services for the elderly in rural Tanzania? A cross-sectional study // *Global Health Action*. 2020. № 1. P. 184–196.
2. Kačur J., Flegner P., Durdán M., Laciak M. Model predictive control of UCG: An experiment and simulation study // *Information Technology and Control*. 2019. № 4. P. 557–578.
3. Karakaplan M. U., Kutlu L., Tsionas M. G. A solution to log of dependent variables with negative observations // *Journal of Productivity Analysis*. 2020. № 2-3. P. 107–119.
4. Kuzmin P., Kalashnikov V., Kalashnykova N., Watada J. The Great Depression: Econometric Analysis and Fuzzy Regression // *Journal of Advanced Computational Intelligence and Intelligent Informatics*. 2020. № 6. P. 785–791.

5. Volchik V., Maslyukova E., Strielkowski W. Spatial effects of informal employment and precarization in urban centres and regions // *Theoretical and Empirical Researches in Urban Management*. 2020. № 4. P. 21–37.
6. Носков С. И. Технология моделирования объектов с нестабильным функционированием и неопределенностью в данных. Иркутск : Облформпечать, 1996. 320 с.
7. Носков С. И. Метод максимальной согласованности в регрессионном анализе // *Известия Тульского государственного университета. Технические науки*. 2021. № 10. С. 380–385.
8. Носков С. И. Реализация конкурса регрессионных моделей с применением критерия согласованности поведения // *Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии*. 2021. № 2. С. 153–160.
9. Васильев С. Н., Селедкин А. П. Синтез функции эффективности в многокритериальных задачах принятия решений // *Известия АН СССР. Техническая кибернетика*. 1980. № 3. С. 186–190.
10. Носков С. И., Лакеев А. В. РС-решения и квазирешения интервальной системы линейных алгебраических уравнений // *Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления*. 2021. № 3. С. 262–276.

References

1. Tungu M., Amani P. J., Hurtig A.-K. [et al.]. Does health insurance contribute to improved utilization of health care services for the elderly in rural Tanzania? A cross-sectional study. *Global Health Action*. 2020;(1):184–196.
2. Kačur J., Flegner P., Durdán M., Laciak M. Model predictive control of UCG: An experiment and simulation study. *Information Technology and Control*. 2019;(4):557–578.
3. Karakaplan M.U., Kutlu L., Tsionas M.G. A solution to log of dependent variables with negative observations. *Journal of Productivity Analysis*. 2020;(2-3):107–119.
4. Kuzmin P., Kalashnikov V., Kalashnykova N., Watada J. The Great Depression: Econometric Analysis and Fuzzy Regression. *Journal of Advanced Computational Intelligence and Intelligent Informatics*. 2020;(6):785–791.
5. Volchik V., Maslyukova E., Strielkowski W. Spatial effects of informal employment and precarization in urban centres and regions. *Theoretical and Empirical Researches in Urban Management*. 2020;(4):21–37.
6. Noskov S.I. *Tekhnologiya modelirovaniya ob"ektov s nestabil'nym funktsionirovaniem i neopredelennost'yu v dannykh = Technology of modeling objects with unstable functioning and uncertainty in data*. Irkutsk: Oblinformpechat', 1996:320. (In Russ.)
7. Noskov S.I. The method of maximum consistency in regression analysis. *Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki = Proceedings of Tula State University. Technical sciences*. 2021;(10):380–385. (In Russ.)
8. Noskov S.I. Implementation of the competition of regression models using the criterion of consistency of behavior. *Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriya: Sistemnyy analiz i informatsionnye tekhnologii = Bulletin of the Voronezh State University. Series: System Analysis and Information Technologies*. 2021;(2): 153–160. (In Russ.)
9. Vasil'ev S.N., Seledkin A.P. Synthesis of the efficiency function in multi-criteria decision-making tasks. *Izvestiya AN SSSR. Tekhnicheskaya kibernetika = Proceedings of the USSR Academy of Sciences. Technical cybernetics*. 1980;(3):186–190. (In Russ.)
10. Noskov S.I., Lakeev A.V. RS-solutions and quasi-solutions of an interval system of linear algebraic equations. *Vestnik Sankt-Peterburgskogo universiteta. Prikladnaya matematika. Informatika. Protsessy upravleniya = Bulletin of the St. Petersburg University. Applied mathematics. Computer science. Management processes*. 2021;(3): 262–276. (In Russ.)

Информация об авторах / Information about the authors

Сергей Иванович Носков

доктор технических наук, профессор,
профессор кафедры информационных
систем и защиты информации,
Иркутский государственный
университет путей сообщения
(Россия, г. Иркутск,
ул. Чернышевского, 15)
E-mail: sergey.noskov.57@mail.ru

Sergey I. Noskov

Doctor of technical sciences, professor,
professor of the sub-department
information systems
and information protection,
Irkutsk State Transport University
(15 Chernyshevskogo street,
Irkutsk Russia)

**Авторы заявляют об отсутствии конфликта интересов /
The authors declare no conflicts of interests.**

Поступила в редакцию/Received 11.01.2022

Поступила после рецензирования/Revised 11.02.2022

Принята к публикации/Accepted 04.03.2022