УДК 51:53

Теоретические и эмпирические модели процессов и их приложения

Н.М. Оскорбин, С.И. Суханов

Алтайский государственный университет (Барнаул, Россия)

Theoretical and Empirical Models of Processes and Their Application

N.M. Oskorbin, S.I. Sukhanov

Altai State University (Barnaul, Russia)

Акцентируется внимание на корректности математического моделирования процессов вне зависимости от выбранных процедур обработки результатов наблюдений. Описываются теоретические модели процессов, исходные предположения для которых априори выполняются, и эмпирические модели, исходные базы данных и знаний которых могут быть недостоверными. Рассматриваются прикладные задачи прогноза и оценки параметров, обоснования оптимальных решений, которые базируются на теоретических и эмпирических моделях процессов. Предлагаются методы конструирования и оценки работоспособности модельных решений, которые зависят от уровня неопределенности описания процессов исследуемого объекта. Исходя из вышесказанного, проводится анализ теоретических моделей и прикладных задач, решаемых с их использованием. Также рассматриваются методы анализа достоверности исходных предположений и способов их контроля, схемы обработки информации: задача прогноза; задача точечной оценки; задача обоснования оптимального решения. В заключении описываются некоторые приемы построения эмпирических моделей процессов, основанные на интервальных оценках.

Ключевые слова: регрессионный анализ, прикладной интервальный анализ, модели прогноза, модели оценки параметров, уровень неопределенности.

In this study, we focus on the accuracy of mathematical modeling of processes, regardless of selected observation results processing routines. This paper presents theoretical models of processes with assumptions fulfilled a priori and empirical models with unreliable original database and knowledge base. We consider applied problems of forecast and evaluation of target parameters, provisioning of grounds for optimal solutions based on theoretical and empirical models of processes. We introduce methods to design and evaluate the efficiency of model solutions which depend on the level of uncertainty while describing the processes of the object under study. Based on that, theoretical models and applied problems solved by their use are analyzed. Also, we discuss methods of reliability assumptions analysis and their control and techniques of information processing: the forecasting problem, the point estimation problem, the optimal solutions problem. In conclusion, some techniques based on interval estimates for developing empirical models of processes are presented.

Key words: regression analysis, applied interval analysis, forecast models, parameter evaluation models, level of uncertainty.

DOI 10.14258/izvasu(2017)1-21

Введение. В работе рассматриваются новые аспекты обработки наблюдений при построении функциональных зависимостей. Краткая формулировка задачи: по таблице наблюдений $(x_j, y_j)_{j=1}^N$ за входными переменными $x \in X \subset R^n$ и выходной переменной $y \in R$ моделируемого процесса оценить значение вектора a заданной функции $y = f(a, x), x \in \overline{X}$. Различные аспекты этой задачи и методы ее решения рассматривались, например, в работах [1–11].

Начиная с Гаусса задача решается методом наименьших квадратов (МНК), который в случае линейной по параметрам функции f сводится к системе линейных алгебраических уравнений. В начале прошлого века с развитием математической статистики появилась дополнительная возможность количественной оценки точности вычисленных параметров и значений функции. Фундаментальными научными результатами этого развития являются доказательство оптимальности оценок МНК в классе линейных операторов математической статистики, разработка процедур проверки статистических гипотез в четко сформулированных исходных предположениях [1].

Во второй половине прошлого века эти методы оформились как регрессионный и корреляционный анализы [3, 4]. В это же время накапливаются проблемы их применения на практике. Критика ведется по выполнимости исходных предположений по двум главным направлениям: ограничения в использовании вероятностно-статистических методов на практике [12–14] и по путям «выхода из кризиса» при возможных нарушениях исходных предположений [3–5].

Характерно для науки, что в это время был предложен и развит новый метод решения рассматриваемой задачи, который основан на интервальной математике [2, 6–10].

В ряде работ ранее и сейчас прикладной интервальный анализ данных рассматривается как альтернативный подход к традиционному, что, в частности, выражается использованием термина «нестатистический» [6, 10]. В этой связи следует отметить сравнительные исследования статистических и интервальных оценок, выполненные в работах [15, 10], которые исключили их доминирование: в зависимости от условий моделирования интервальные оценки могут быть лучше или хуже оценок МНК. Сблизило эти методы и то, что условия их применения могут одинаково не выполняться и трудно поддаются контролю, особенно в слабоструктурированных приложениях.

Необходимым условием прикладной ценности математических моделей является их достоверность. Это требование выражается в том, что для истинного процесса $y = f_u(a^*, x)$ рассматриваемого объекта O выполнено условие $f_u(x) \in F(x) = [F_H(x), F_V(x)]\ \forall x \in X$. Здесь множество $F(x)$ выступает интервалом неопределенности полученной модели процесса $y = f(a, x), x \in X$.

Погрешности моделирования можно выразить множеством $A(x)$ оценок вектора параметров a. В регрессионном анализе это множество определяется уровнем доверительной вероятности расчетной функции распределения вектора параметров a, в интервальном анализе $A(x)$ является решением интервальных систем уравнений. Множества $F(x)$ и $A(x)$ согласованы, т.е. $F_H(x) = \max\limits_{a \in A(x)} f(a,x);\quad F_V(x) = \min\limits_{a \in A(x)} f(a,x), \forall x \in \overline{X}$. Таким образом, теоретическими мы называем модели, для которых включение $f_u(x) \in F(x)$, или $a^* \in A(x)$ обеспечивается априори. Для эмпирических моделей необходимы дополнительные процедуры проверки выполнимости этого условия.

Таким образом, акцент на построение теоретических моделей и на корректное построение эмпирических моделей позволяет повысить их прикладную ценность. В работе приведен метод двух моделей, который, учитывая особенности прикладных задач, повышает эффективность модельных решений.

Теоретические модели процессов и их приложения. Рассмотрим схемы обработки информации по объектам и процессам, для которых имеется хорошая теоретическая база знаний и экспериментальная практика (рис. 1).

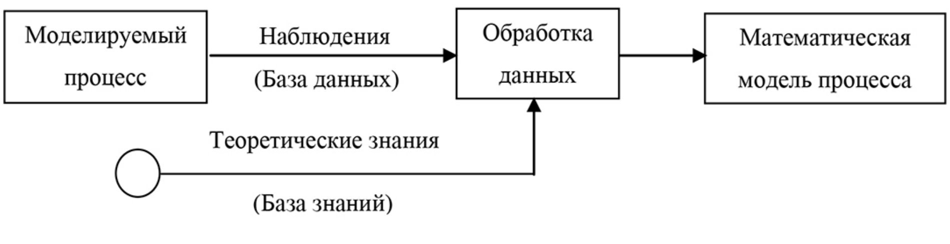

Рис. 1. Схема построения теоретических моделей процессов

Результатом теоретического моделирования известного с точностью до параметров процесса $y = f(a, x), x \in X$ является множество $a \in A(x)$.

В работе [11, с. 13] предлагается расширить формализацию задачи и дополнительно к схеме рисунка 1 провести математическое моделирование процесса использования модели (рис. 2). Двухэтапная схема обработки информации характеризуется тем, что производится построение двух математических моделей, исходной для моделируемого объекта O и прикладной, при конструировании которой учитываются дополнительные условия решаемой прикладной задачи. В рассматриваемом случае используется известный подход исследования операций, при котором обоснование решений осуществляется по второй модели, а оценка их эффективности (работоспособности) выполняется по первой. Заметим, что в частных случаях первая и вторая модели могут совпадать.

Рассмотрим далее задачи обработки информации по схеме, представленной на рисунке 2.

Задача прогноза. Пусть на множестве точек $x \in X_\Pi \subseteq X$ для объекта O необходимо получить оценки $\hat{y}$ выходной переменной y. Данную задачу решаем методами исследования операции при априори заданной произвольной структуре модели прогноза $y_\Pi = F_\Pi(x,b)$, $x \in X_\Pi; b \in R^n$ путем нахождения точечной оценки вектора неизвестных параметров: $b^* \in R^n$.

Рис. 2. Схема обработки информации на стадии использования математической модели

В рамках вероятностно-статистических подходов данная задача впервые поставлена и исследована для частных случаев в работе [5, с. 152–153], в которой модель прогноза настраивается решением следующей задачи:

$$s^2 = \min_{b \in R^n} M_{x,a}(f(a,x) - f_П(b,x))^2, \qquad (1)$$

где квадрат разности прогнозной модели и описания объекта O усредняется по переменным a и $x \in X_П \subseteq X$ в предположении, что для пользователей известно вероятностное распределение входных переменных на интервале времени использования прогнозной модели.

Указанный способ конструирования моделей прогноза можно использовать в различных исходных условиях построения теоретических моделей процессов, в том числе и отличных от предположений регрессионного анализа. В последнем случае функция распределения вектора параметров a и оценки множества $A(x)$ вычисляются теоретическими методами или с использованием имитационного моделирования.

Конструирование уравнений прогноза согласно (1) существенно расширяет область применения математических моделей на практике. Для выбора уравнений прогноза и условий его использования нет необходимости использовать истинную модель объекта O, которая может оказаться сложной или не реализуемой по условиям прогноза, например, из-за технических ограничений. Иногда можно сознательно изменить структуру модели прогноза с целью повышения его качества. Так, в [5, с. 153] приведен пример модификации уравнения прогноза, которая приводит к снижению на 14% ошибки прогноза по сравнению с ошибкой при использовании функции $y = f(a, x), x \in X$.

Для оценки работоспособности модели прогноза используется, как и в регрессионном анализе, коэффициент детерминации и рекомендованные для него пороговые значения.

Задача точечной оценки вектора параметров β в рамках вероятностно-статистических подходов. Считаем, что модель исследуемого объекта известна с точностью $\beta \in B(x)$. Лучшее приближение b^* вектора параметров β зависит от условий их использования, задаваемых множеством $x \in X_П \subseteq X$ и от вероятностных характеристик вектора x. Критерий качества оценки связан с расстоянием $\left\|\beta - b^*\right\|$. Задача оценки: найти b^* из условия

$$\Delta = \min_{b \in R^n} M_x \max_{a \in A(x)} \left\|a - b\right\|. \qquad (2)$$

В частном случае, когда множество A не зависит от x, задача (2) сводится к поиску минимального шара, содержащего это множество: $\Delta = \min_{b \in R^n} \max_{a \in A} \left\|a - b\right\|$.

Задача обоснования оптимального решения x^* для выбранного объекта O. Пусть $\varphi(x, w, y)$ — целевая функция задачи выбора и вектор w точно известен ЛПР. Статистически оптимальное решение находим так:

$$\tilde{x}^*(w) \in \arg\max_{x \in X} M_a\left[\varphi(x, w, y)\right]. \qquad (3)$$

Построение специальных моделей обоснования решений, отличных от (3), возникает, например, в следующих случаях: наличие ограничений на реализацию решения $\tilde{x}^*(w)$; необходимость использования простых моделей управления; информационное вырождение оценок вектора a при текущей идентификации модели.

В качестве иллюстрации выбора простой модели принятия решений вместо (3) рассмотрим объект O, у которого $\varphi(x, w, y) = -(y - w)^2$, $y = a\,x$, $x \in R$; w — известный ЛПР параметр. Пусть на стадии описания объекта найден интервал $[A_H, A_V]$ оценок параметра a, и аналогом модели (3) принята минимаксная задача обоснования решений: $\tilde{x}^*(w) \in \arg\max_{x \in X} \min_{a \in [A_H, A_V]} \left[-(a\,x - w)^2\right]$.

Оптимальное решение имеет вид: $\tilde{x}^*(w) = 2w / (A_H + A_V)$. Выберем для обоснования решения x^* детерминированную модель $y = \gamma\,x$, $x \in R$, где γ — параметр упрощенной модели. Оптимальное решение имеет вид: $\tilde{x}^*(w) = w / \gamma$. Если принять значение параметра γ, равное середине интервала $[A_H, A_V]$, то детерминированная задача принятия решений эквивалентна минимаксной задаче для всех значений w.

Развитие данного подхода к выбору альтернативных (3) моделей обоснования оптимальных решений проведено в работе [11], в которой метод двух моделей применяется для спектра условий их применения. При этом подобно задаче синтеза моделей прогноза (1) первая модель является рабочей, а вторая используется для настройки параметров первой и оценки ее работоспособности.

Моделирование процессов: прикладной интервальный анализ. Метод двух моделей может найти применение в прикладном интервальном анализе, который рассматривался, например, в работах [8, 12, 16, 17]. Даже в случае линейной по параметрам модели процесса множество неопределенности может зависеть от значений входных переменных. Эта зависимость возникает, например, при использовании в качестве априорной информации функциональных неравенств. Так, еще в работе [2, с. 702] отмечается, что в ряде случаев существенное использование в численных алгоритмах может найти полученная теоретическим путем информация о расположении и свойствах решения (границы самого решения и его производных и т.п.).

Рассмотрим случай построения линейной модели процесса по таблице наблюдений, в которой переменные $(x_1, x_2, ..., x_n)$ измерены безошибочно, а y_j заданы интервалом $Y_j = [y_j^H, y_j^V], j = 1,...,N$, где

$y_j^H = y_j^R - \varepsilon_j^0$; $y_j^V = y_j^R + \varepsilon_j^0$; y_j^R – результат, а ε_j^0 — оценка ошибки измерения *j–го* опыта. Тогда оценки $A(N)$ истинных коэффициентов модели $y_0 = a_1x_1 + ... + a_nx_n$ удовлетворяют системе N двухсторонних неравенств, которые соответствуют интервальным системам линейных алгебраических уравнений (ИСЛАУ):

$$y_j^R - \varepsilon_j^0 \le a_1x_{1j} + ... + a_nx_{nj} \le y_j^R + \varepsilon_j^0; \quad j = 1,...,N. \tag{4}$$

Множество $A(N)$ в прикладном интервальном анализе играет фундаментальную роль. Во-первых, оно содержит всю полученную информацию об искомой зависимости, т.е. является теоретическим результатом ее оценивания (первая модель процесса). Кроме того, прикладные модели (модели 2-го этапа) конструируются с использованием этого множества, в том числе и для проверки работоспособности полученных моделей.

В литературе множество $A(N)$ носит название объединенного множества решений ИСЛАУ (4) [16]. Как пример решения проблем пользователей приведем интервальную оценку $[y^H(x^p), y^V(x^p)]$ прогноза выходной переменной моделируемого процесса в заданной точке $x^p \in X_П$:

$$\begin{aligned} y^H(x^p) &= \min_{a \in A(N)} (a_1x_1^p + ... + a_nx_n^p); \\ y^V(x^p) &= \max_{a \in A(N)} (a_1x_1^p + ... + a_nx_n^p). \end{aligned} \tag{5}$$

Есть основания утверждать, что примерами пользовательских задач оценки параметров и обоснования решений для рассматриваемого случая выступают допусковое и управляемое множества решений ИСЛАУ [8, 16].

Эмпирическое моделирование процессов. Рассматриваем моделирование процессов, для которых исходные базы знаний, базы данных могут быть недостоверными. В настоящее время выделено два направления в задачах моделирования процессов: развитие методов машинного поиска скрытых закономерностей баз данных и знаний [17–20] и разработка специальных эвристических процедур контроля выполнимости исходных предположений [13, 14, 11]. Далее мы рассматриваем только второе направление.

В работах Ю.И. Алимова содержатся рекомендации по совершенствованию вероятностно-статистического моделирования, суть которых состоит в увеличении числа N наблюдений за объектами-аналогами моделируемого процесса и формировании групп наблюдений для проверки воспроизводимости статистических характеристик. Этот прием может позволить либо повысить доверие аналитиков к эмпирической модели, либо признать задачу моделирования нереализуемой. По нашему мнению, критика этих рекомендаций, например, в работе [12] не является справедливой. Более того, мы считаем возможным развитие этого подхода как теоретической основы эмпирического моделирования процессов для проверки выполнимости утверждения «так было, следовательно, так будет».

Применительно к эмпирическим моделям интервального оценивания некоторые результаты представлены в работе [11] и оформлены в рамках метода центра неопределенности (МЦН) [21, 22]. Эмпирические модели, доверие к которым является приемлемым, могут, как и теоретические модели, выступать основой решения прикладных задач.

Заключение. Мы рассмотрели методы конструирования прикладных моделей прогноза, оценки параметров, обоснования оптимальных решений, которые базируются на исходных теоретических и эмпирических моделях процессов. В ходе исследования проведен анализ теоретических моделей и прикладных задач, решаемых с их использованием. Описывается подход построения эмпирических моделей процессов, который включает эвристические методы анализа достоверности исходных предположений.

Библиографический список

1. Гмурман В.Е. Теория вероятностей и математическая статистика. — М., 2011.

2. Канторович Л.В. О некоторых новых подходах к вычислительным методам и обработке наблюдений // Сибирский математический журнал. — 1962. — Т. 3, № 5.

3. Химмельблау Д. Анализ процессов статистическими методами. — М., 1973.

4. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. — М., 1973.

5. Люблинский Р.Н., Оскорбин Н.М. Методы декомпозиции при оптимальном управлении непрерывным производством. — Томск, 1979.

6. Вощинин А.П., Бочков А.Ф., Сотиров Г.Р. Интервальный анализ данных как альтернатива регрессионному анализу // ЗЛ. — 1990. — № 7.

7. Спивак С.И. Детальный анализ применения методов линейного программирования при определении параметров кинетической модели // Математические проблемы химии. — Новосибирск, 1975..

8. Шарый С.П. Решение интервальной линейной задачи о допусках // Автоматика и телемеханика. — 2004. — № 10.

9. Оскорбин Н.М., Жилин С.И., Максимов А.В. Построение и анализ эмпирических зависимостей методом центра неопределенности // Известия Алтайского гос. ун-та. — 1998. — № 1.

10. Суханов В.А. Исследование эмпирических зависимостей: нестатистический подход. — Барнаул, 2007.

11. Максимов А.В., Оскорбин Н.М. Многопользовательские информационные системы: основы теории и методы исследования. — 2-е изд., испр. и доп. — Барнаул, 2013.

12. Вощинин А.П. Интервальный анализ данных: развитие и перспективы // ЗЛ. — 2002. — Т. 68, № 1.

13. Тутубалин В.Н. Вероятность, компьютеры и обработка результатов экспериментов // УФН. — 1993. — Т. 163, № 7.

14. Алимов Ю.И., Кравцов Ю.А. Является ли вероятность «нормальной» физической величиной? // УФН. — 1992. — Т. 162, № 7.

15. Оскорбин Н.М., Жилин С.И., Дронов С.В. Сравнение статистической и нестатистической оценок параметров эмпирической зависимости // Известия Алтайского гос. ун-та. — 1998. — № 4.

16. Шарый С.П. Конечномерный интервальный анализ. — Новосибирск, 2016.

17. Загоруйко Н.Г. Прикладные методы анализа данных и знаний. — Новосибирск, 1999.

18. Подружко А.А., Подружко А.С. Интервальное представление полиномиальных регрессий. — М., 2003.

19. Заде Л.А. Понятие лингвистической переменной и его применение к понятию приближенного решения. — М., 1976.

20. Пономарев И.В., Славский В.В. Нечеткая модель линейной регрессии // Доклады академии наук. — 2009. — Т. 428, № 5.

21. Мадияров М.Н., Оскорбин Н.М. Применение интервальной математики в задачах регрессионного анализа // Постиндустриальный мир: зеленый рост и зеленая экономика : сборник материалов республиканской науч.-практ. конф. — Усть-Каменогорск, 2016.

22. Суханов С.И. Интервальный анализ в задачах моделирования пространственного положения геообъектов. — Барнаул, 2016.