

Примеры интервального анализа данных в задачах моделирования процессов

М.Н. Мадияров¹, Н.М. Оскорбин², С.И. Суханов²

¹Восточно-Казахстанский государственный университет им. С. Аманжолова (Усть-Каменогорск, Казахстан)

²Алтайский государственный университет (Барнаул, Россия)

Examples of Data Interval Analysis in Process Modeling Problems

M.N. Madiyarov¹, N.M. Oskorbin², S.I. Sukhanov²

¹Sarsen Amanzholov East Kazakhstan State University (Ust-Kamenogorsk, Kazakhstan)

²Altai State University (Barnaul, Russia)

Проводится исследование возможностей применения теории интервальных систем линейных алгебраических уравнений (ИСЛАУ) для решения задач математического моделирования процессов по экспериментальным данным. Предполагается, что моделируемый процесс описывается выходной переменной и совокупностью входных переменных, детерминированная связь которых описывается линейным уравнением. При моделировании процесса используется теоретический подход, при котором исходные предположения относительно структуры модели и границы интервалов ошибок измерения всех переменных являются достоверными и не требуют проверки их выполнимости методами разведочного анализа. Указанная постановка задачи моделирования процессов позволяет сосредоточить внимание на исследовании прикладной ценности основных множеств решений ИСЛАУ: объединенного, допускового и управляемого. В целях достижения наглядности методов и полученных результатов исследование проводится на конкретном примере процесса, для которого параллельно можно проводить визуальный анализ данных.

Ключевые слова: моделирование процессов, интервальные системы линейных алгебраических уравнений, модели прогноза, модели оценки параметров.

DOI 10.14258/izvasu(2018)1-20

Введение

В работе проводится исследование возможностей применения теории интервальных систем линейных алгебраических уравнений (ИСЛАУ) для решения задач математического моделирования процессов по экспериментальным данным. В данном случае считается, что внутренний шум в описании объекта

In this paper, we study the possibilities of applying the theory of interval systems of linear algebraic equations to solve problems of mathematical modeling of processes using experimental data. It is assumed that the simulated process is described by the output variable and a set of input variables with deterministic connections described by a linear equation. During simulation, a theoretical approach is used for initial assumptions about the structure of the model and the boundaries of measurement error intervals for all variables being reliable and not requiring verification of their feasibility by exploratory analysis methods. This formulation of the process modeling problem allows focusing on the study of practical values of the basic sets of solutions of interval systems of linear algebraic equations: combined, tolerated and controlled. For the sake of clarity of the methods and the results obtained, the study is conducted on the specific example of the process with a simultaneous visual analysis of data being carried out.

Key words: modeling processes, interval systems of linear algebraic equations, forecast models, parameter evaluation models.

отсутствует, а неопределенности таблицы наблюдений связаны с ошибками измерения входных и выходной переменных, которые принадлежат известным симметричным относительно нулевого значения интервалам. Различные аспекты этой задачи и нестатистические методы ее решения рассматривались, например, в работах [1–11].

При записи ИСЛАУ мы, следуя [6], используем неформальный международный стандарт, согласно которому интервалы и интервальные объекты обозначаются жирным шрифтом. В целях достижения наглядности методов и полученных результатов исследование проводится на примере процесса планирования прибыли корпорации, для которого параллельно можно проводить визуальный анализ данных.

Считаем, что корпорация состоит из n предприятий и заинтересована в получении прибыли. Оценка величины прибыли на очередной период времени (месяц, квартал) проводится на основе планов ее получения каждым предприятием. Управляющая компания объединения не производит продукцию и не осуществляет финансовые (инвестиционные) услуги. В случае такой возможности в корпорации создается самостоятельная компания, которая включена в состав рассматриваемых предприятий. Пусть A_i^{N+1} , $i=1, \dots, n$ — оценки прибыли предприятий и B^{N+1} — расчетное значение прибыли корпорации, полученное аналитиками на плановый период времени. Предполагаем, что управляющая компания располагает статистикой плановых оценок прибыли предприятий и корпорации за N временных периодов в прошлом:

$$(A_1^j, A_2^j, \dots, A_n^j, B^j), \quad j = 1, \dots, N. \quad (1)$$

Кроме того, полагаем выполнение балансового уравнения по прибыли корпорации в каждом периоде наблюдения, которое запишем в следующем виде. Пусть A_i' , $i=1, \dots, n$ — значения истинных величин прибыли предприятий, которые априори неизвестны как предприятиям, так и управляющей компании. Тогда действительная прибыль корпорации B' равна сумме прибылей предприятий (налоги и транзакции при переходе прав собственности на прибыль не учитываем): $B' = \sum_{i=1}^n A_i'$.

По результатам финансового анализа управляющая компания располагает достоверными оценками модулей предельных ошибок, которые имели место в прошлом при обосновании данных таблицы (1). Эти оценки обозначим так:

$$(\varepsilon_1^j, \varepsilon_2^j, \dots, \varepsilon_n^j, \varepsilon_B^j), \quad j = 1, \dots, N. \quad (2)$$

Экономическая ситуация и существующие технологии финансового планирования корпорации позволяют ее аналитикам обосновать оценку будущей прибыли корпорации B^{N+1} , оценить границы абсолютных значений погрешностей планирования прибыли предприятий и корпорации в целом:

$$(\varepsilon_1^{N+1}, \varepsilon_2^{N+1}, \dots, \varepsilon_n^{N+1}, \varepsilon_B^{N+1}). \quad (3)$$

Далее учитывается, что приведенные данные соответствуют правильным наблюдениям [7] (например, не содержат выбросов). Методами прикладного анализа данных предлагается выполнить следующие аналитические исследования:

- Получить интервальную оценку \mathbf{B}^{N+1} прибыли корпорации при существующей технологии финансового планирования, которая является достоверной, т.е. содержит истинное значение будущей прибыли корпорации B^{N+1} .

- Проверить согласованность существующей технологии расчета прибыли корпорации B^{N+1} его интервальной оценке, т.е. проверить условие $B^{N+1} \in \mathbf{B}^{N+1}$.

- Качественно оценить эффективность решения задач математического моделирования процессов с использованием множеств решений ИСЛАУ: объединенного, допускового и управляемого.

Исследование выполнялось методами вычислительного эксперимента.

Математические методы и программные средства исследования

Прикладной интервальный анализ в нашем случае линейных моделей детерминированных процессов проводится с использованием множеств решений ИСЛАУ, коэффициенты и правая часть которой записаны по результатам интервальных наблюдений. В матричной форме ИСЛАУ записывается интервальной $N \times n$ матрицей коэффициентов и $N \times 1$ интервальным вектором правой части в следующем виде:

$$\mathbf{Ax} = \mathbf{b}. \quad (4)$$

Интервальное задание СЛАУ (4) понимается так, что элементы матриц \mathbf{A}, \mathbf{b} заданы интервалами: $A^H \leq \mathbf{A} \leq A^V$ и $B^H \leq \mathbf{b} \leq B^V$.

Применительно к задачам анализа данных в литературе [2, 3, 5, 6] рассматриваются три базовых множества решений ИСЛАУ: объединенное, допусковое, управляемое. Предикатная формула объединенного множества решений $\Xi_{uni}(\mathbf{A}, \mathbf{b})$ записывается так:

$$\Xi_{uni}(\mathbf{A}, \mathbf{b}) = \{x \in R^n / (\exists A \in \mathbf{A})(\exists B \in \mathbf{b})(Ax = B)\}. \quad (5)$$

В работе [3, с. 323] утверждается, что «вычисление для объединенного множества решений внешних по координатных оценок с любой заданной абсолютной или относительной точностью есть NP-трудная задача». В частном случае положительных компонент решения ИСЛАУ объединенное множество решений задается системой линейных неравенств, которые, следуя [5, с. 112], запишем ниже.

Допусковое $\Xi_{doi}(\mathbf{A}, \mathbf{b})$ и управляемое $\Xi_{cti}(\mathbf{A}, \mathbf{b})$ множества решений определяются условиями принадлежности интервалам:

$$\Xi_{\text{int}}(\mathbf{A}, \mathbf{b}) = \{x \in R^n / (\mathbf{Ax} \subseteq \mathbf{b})\}; \quad \Xi_{\text{ext}}(\mathbf{A}, \mathbf{b}) = \{x \in R^n / \mathbf{Ax} \supseteq \mathbf{b}\}. \quad (6)$$

В прикладном интервальном анализе множества (6) допустимо исследовать методами линейного программирования, а для возможности использования стандартного симплекс метода вектор $x \in R^n$ можно представить в виде разности двух неотрицательных векторов x', x'' : $x = x' - x''$. Если априори известно,

что множества лежат в заданном квадранте декартовой системы координат R^n , то заменой переменных можно упростить соответствующие задачи ЛП. Так, для R_+^n положим $x'' = 0$. Тогда множества (5), (6) задаются в R_+^n системами неравенств соответственно:

$$\begin{cases} A^V x \geq B^V \\ A^H x \leq B^V \end{cases}; \quad \begin{cases} A^V x \leq B^V \\ A^H x \geq B^H \end{cases}; \quad \begin{cases} A^V x \geq B^V \\ A^H x \leq B^H \end{cases}. \quad (7)$$

Запишем задачи исследования множеств решений ИСЛАУ, которые возникают при моделировании процессов. Обозначим $X_S \subset R^n$ — одно из трех введенных множеств решения ИСЛАУ в задаче анализа экспериментальных данных. Эти задачи запишем в обозначениях и предположениях приведенного примера.

$$\hat{B}^{H,N+1} = \min_{x \in X_S} (A_1^{H,N+1} x_1 + \dots + A_n^{H,N+1} x_n); \quad \hat{B}^{V,N+1} = \max_{x \in X_S} (A_1^{V,N+1} x_1^p + \dots + A_n^{V,N+1} x_n^p). \quad (8)$$

Заметим, что прогноз по измеренным значениям A_i^{N+1} , $i = 1, \dots, n$ в (8) не является корректным, поскольку оцениваемый интервал может не содержать истинного значения B^{N+1} прогнозируемой прибыли.

2. Задача оценки коэффициентов линейной зависимости моделируемого процесса.

Для нашего примера предполагается, что коэффициенты баланса прибыли корпорации неизвестны. В качестве интервальной оценки этих коэффициентов можно решить $2n$ задач линейного программирования. Например, x_1 принадлежит интервалу $[x_1^H, x_1^V]$:

$$x_1^H = \min_{x \in X_S} x_1; \quad x_1^V = \max_{x \in X_S} x_1. \quad (9)$$

Дополнительно запишем задачу принадлежности любой заданной точки $x^0 \in R^n$ заданному множеству. Пусть, например, $x^d \in R^n$ — истинное значение параметров моделируемого детерминированного процесса, а $X_S \subset R^n$ — одно из трех введенных множеств решения ИСЛАУ в задаче анализа экспериментальных данных. Точка x^d принадлежит множеству $X \subset R$ тогда и только тогда, когда $\delta_S = 0$, где δ_S — решение следующей задачи квадратичного программирования:

$$\delta_S = \min_{x \in X_S} \|x^d - x\|. \quad (10)$$

Имитационное моделирование задачи оценки прибыли корпорации

Имитация условий анализа прибыли корпорации проводилась в среде Excel при следующих исходных данных: $n = 3$; $N = 12$; истинные значения прибылей A_i^t , $i = 1, \dots, 3$ в каждом из $N + 1$ временных перио-

1. Задача прогноза выходной переменной моделируемого процесса на период времени $N + 1$.

Учитываем при прогнозе только оценки прибыли предприятий: $A_i^{H,N+1}$; $A_i^{V,N+1}$, $i = 1, \dots, n$, тогда:

дов принимались как равномерно распределенные псевдослучайные числа на интервалах $[0, 100]$; соответствующие значения прибыли корпорации B^t определялись на основе балансного уравнения; ошибки измерения (2) и (3) во всех испытаниях принимались равномерно распределенными на одинаковых симметричных относительно нуля интервалах: $[-5, 5] \times [-5, 5] \times [-5, 5] \times [-\varepsilon_B, \varepsilon_B]$. Здесь верхние оценки ошибки выходной переменной задавались переменными для обеспечения условия, при котором исследуемое множество решений ИСЛАУ не пусто и сравнимо с двумя другими. Пример точных значений прибылей корпорации и их оценок представлен в таблицах 1 и 2.

Таблица 1
Данные без ошибок измерения

	A_1^t	A_2^t	A_3^t	B^t
1	23,93	19,84	97,14	140,91
2	13,84	91,08	65,11	170,03
3	35,27	18,48	70,67	124,42
4	68,21	39,87	84,02	192,09
5	76,78	22,20	13,35	112,33
6	24,28	88,86	64,10	177,23
7	11,60	6,28	26,29	99,18
8	12,92	49,01	83,69	145,62
9	91,99	26,19	21,89	140,07
10	29,43	61,70	16,77	107,90
11	58,73	80,64	95,76	235,12
12	83,98	89,17	5,45	178,60
13	17,01	73,08	51,91	142,00

Таблица 2

Данные с ошибками измерения, при $\varepsilon_B = 10$

	A'_3	A'_3	A'_3	B'
1	26,99	19,99	98,11	149,74
2	17,80	93,8	68,41	160,7
3	38,41	14,64	67,47	127,32
4	67,48	38,84	88,28	191,76
5	75,53	23,05	11,26	113,07
6	19,37	83,87	64,64	182,70
7	16,57	57,07	30,77	94,54
8	17,00	47,84	79,19	150,69
9	89,93	27,74	22,38	147,71
10	24,47	65,00	18,49	107,10
11	57,86	84,12	95,31	240,01
12	79,54	89,72	4,98	169,15
13	15,94	73,27	47,94	135,37

Объединенное множество решений ИСЛАУ содержит точку $x^d = (1, 1, 1)$ по определению этого множества для правильных измерений, что следует и по результатам численного решения задачи (10).

Для данных таблицы 2 решение задач (8) является следующим:

$$\hat{B}^{H,N+1} = 109,7; \hat{B}^{V,N+1} = 160,3; B^{C,N+1} = 0.5(\hat{B}^{H,N+1} + \hat{B}^{V,N+1}) = 135; \Delta = 17.8\%. \quad (11)$$

Для сравнения приведем значения и погрешности прогноза по оценкам аналитиков. Эти данные в порядке выражения (11) имеют следующие значения: 125,4; 145,4; 135,4; 7,0%. Приведем соответствующие значения, полученные с использованием балансового уравнения по нижним и верхним значениям прибылей предприятий: 122,2; 152,2; 137,2; 10,6%. Приведенные числовые данные не противоречат визуальному анализу и свойствам объединенного мно-

жества решений. Следует отметить, что использование объединенного множества решений ИСЛАУ в качестве инструмента прогноза прибыли корпорации в рассмотренном случае не позволяет улучшить оценки, полученные визуальным анализом данных.

Оценки параметров балансового уравнения в предположении, что точка $x^d = (1, 1, 1)$ аналитикам неизвестна, получим решением задач (9) (табл. 3).

Таблица 3

Оценки объединенного множества решений ИСЛАУ

Оценки $\Xi_{uni}(\mathbf{A}, \mathbf{b})$	Индекс параметров		
	$i = 1$	$i = 2$	$i = 3$
Нижнее значение	0,77	0,57	0,72
Верхнее значение	1,45	1,26	1,39
Погрешность, %	34,1%	34,2%	33,5%

Допусковое множество решений ИСЛАУ для данных таблицы 2 и заданных предельных значений погрешностей измерения является пустым. Этот результат вполне согласуется с исследованиями С.П. Шарого [3], т.е. в данном случае произведение \mathbf{Ax} получает «большой размах» в сравнении с размахом вектора \mathbf{b} . В нашем случае для данных таблицы 1 и принятых оценок погрешностей измерения, в которых $\varepsilon_B = 20$, с дополнительным коэффициентом расширения $k_B = 2$ размаха вектора \mathbf{b} получена ИСЛАУ с подхо-

дящими свойствами. Ее исследование для 15 вариантов выборок независимых ошибок измерения показало, что в пяти вариантах допусковое множество не пусто и содержит точку $x^d = (1, 1, 1)$; в восьми вариантах оно не пусто, но необходимое для корректности модели условие $x^d \in \Xi_{tol}(\mathbf{A}, \mathbf{b})$ не выполнено; в двух вариантах множество $\Xi_{tol}(\mathbf{A}, \mathbf{b})$ оказалось пустым.

Для одного из первой группы вариантов приведем решение задач (8):

$$\hat{B}^{H,N+1} = 121,1; \hat{B}^{V,N+1} = 162,5; B^{C,N+1} = 0.5(\hat{B}^{H,N+1} + \hat{B}^{V,N+1}) = 141,8; \Delta = 14.6\%. \quad (12)$$

Для сравнения, как и ранее, приведем анализ погрешностей прогноза оценок параметров балансового уравнения (табл. 4), аналогичный результатам для $\Xi_{int}(\mathbf{A}, \mathbf{b})$. Эти данные имеют следующие зна-

чения: оценки аналитиков — 124,8; 144,8; 134,8; 7,0% при $\varepsilon_B = 10$; оценки по балансу — 122,4; 152,4; 137,4; 10,6%.

Таблица 4

Оценки допускового множества решений ИСЛАУ

Оценки $\Xi_{int}(\mathbf{A}, \mathbf{b})$	Индекс параметров		
	$i = 1$	$i = 2$	$i = 3$
Нижнее значение	0,85	1,00	0,93
Верхнее значение	1,03	1,16	1,07
Погрешность, %	8,2%	8,7%	6,9%

Сравнение данных таблиц 3 и 4 показывает безусловную эффективность процедур моделирования процессов с использованием допускового множества решений. Данный вывод согласуется с результатами С.П. Шарого [2, 6] и полностью оправдывает метод сильного согласования. Однако следует учитывать, что в ряде задач идентификации процессов необходимо выполнение условия принадлежности истинных значений параметров процесса допусковому множеству решений. Заметим, что прямая проверка этого включения встречает существенные трудности. Сходная проблема возникает при оценке информационного множества в задачах моделирования процессов при неправильных наблюдениях [7, с. 53].

Рассмотрим формально по аналогии с вышеизложенным свойства управляемого множества решений ИСЛАУ для данных таблицы 2 и заданных предель-

ных значений погрешностей. Следует отметить, что примеры использования множества $\Xi_{ct}(\mathbf{A}, \mathbf{b})$ в задачах моделирования процессов авторам неизвестны.

Как и следовало ожидать, это множество для рассматриваемой таблицы измерений является пустым. Поступая зеркально схеме исследования множества $\Xi_{int}(\mathbf{A}, \mathbf{b})$, зададим $\varepsilon_B = 5$ с дополнительным коэффициентом сжатия размаха вектора \mathbf{b} , равным 0,6 ($k_B = 0,6$), и получим ИСЛАУ с подходящими свойствами. Ее исследование для 15 вариантов выборок с независимыми ошибками измерений показало, что в 11 вариантах управляемое множество не пусто и содержит точку $(1, 1, 1)$; в четырех вариантах оно не пусто, но необходимое для корректности модели условие $x^d \in \Xi_{ct}(\mathbf{A}, \mathbf{b})$ не выполнено.

Для одного из первой группы вариантов приведем решение задач (8):

$$\hat{B}^{H,N+1} = 119,1; \hat{B}^{V,N+1} = 160,9; B^{C,N+1} = 0.5(\hat{B}^{H,N+1} + \hat{B}^{V,N+1}) = 140; \Delta = 14.7\%. \quad (13)$$

Как видим, эти данные с учетом точности их вычисления совпадают с (12), как и оценки погрешностей с использованием визуальных методов прогнозирования.

Равнозначность по эффективности моделирования процессов показывают оценки параметров ба-

лансового уравнения, полученные как характеристика бруса, содержащего управляемое множество решений (таблица 5). Мы считаем различия данных таблиц 4 и 5 несущественными, поскольку задание управляемого множества решений оказалось более свободным.

Таблица 5

Оценки управляемого множества решений ИСЛАУ

Оценки $\Xi_{ct}(\mathbf{A}, \mathbf{b})$	Индекс параметров		
	$i = 1$	$i = 2$	$i = 3$
Нижнее значение	0,87	0,83	0,89
Верхнее значение	1,16	1,17	1,14
Погрешность, %	14,6%	16,8%	12,4%

Следует отметить, что результаты вычислительного эксперимента показывают, что в данном случае полезность итоговых интервальных оценок для практики является низкой в силу как значительных ошибок, так и доминирования результатов визуального анализа. Это обстоятельство вызвано тем, что все погрешности хао-

тичны, а при анализе данных не учитывается априорная информация. Реально для рассматриваемого примера в качестве такой информации могут выступать, во-первых, множество X_0 , содержащее точку x^d , во-вторых, оценки интервалов прибыли корпорации, полученные с учетом прогноза прибыли предприятий.

Выполнено исследование множеств решения ИСЛАУ, в которой интервалы правой части получены на основе балансового уравнения. Как и следовало ожидать, допустимое и управляемое множества решений ИСЛАУ не пусто и одноэлементное, т.е. содержит одну точку $x^d = (1, 1, 1)$. Характеристика объединенного множества решений качественно соответствует таблице 3. Прогнозные значения выходной переменной соответствуют результатам визуального анализа данных. Данный пример не имеет прикладной ценности и носит чисто теоретический характер.

Заключение

Рассмотрены задачи моделирования процессов по экспериментальным данным с интервальными ошибками измерения как входных, так и выходной переменных. При статистическом подходе математическое моделирование процессов выполняется методами конфлюэнтного анализа.

Для процесса без внутренних шумов в условиях правильных наблюдений методами вычислительного

эксперимента проведено качественное сравнение точностей прогнозирования выходной переменной и оценивания коэффициентов линейной связи для трех множеств решений ИСЛАУ: объединенного, допустимого и управляемого.

При решении задачи интервального анализа коэффициентов линейной зависимости подтверждена эффективность использования принципа сильного согласования С.П. Шарого [2, 6] при условии корректной оценки их истинных значений (в статистических методах определяется понятиями несмещенности и состоятельности). В процессе исследования затронуты проблемы интервального анализа зависимых измерений (измерений, в которых ошибки взаимно зависимы), учтена априорная информация при записи ИСЛАУ и поиске ее решения.

Полученные результаты позволяют уточнить методические подходы применения теоретических результатов ИСЛАУ в задачах анализа данных и математического моделирования реальных процессов.

Библиографический список

1. Канторович Л.В. О некоторых новых подходах к вычислительным методам и обработке наблюдений // Сибирский математический журнал. — 1962. — Т. 3, № 5.
2. Шарый С.П. Решение интервальной линейной задачи о допусках // Автоматика и телемеханика. — 2004. — № 10.
3. Шарый С.П. Конечномерный интервальный анализ. — Новосибирск, 2017.
4. Оскорбин Н.М., Жилин С.И., Максимов А.В. Построение и анализ эмпирических зависимостей методом центра неопределенности // Изв. Алт. гос. ун-та. — 1998. — № 1.
5. Жолен Л. Прикладной интервальный анализ. — М.; Ижевск, 2005.
6. Шарый С.П. Сильная согласованность в задаче восстановления зависимостей при интервальной неопределенности данных // Вычислительные технологии. — 2017. — Т. 22, № 2.
7. Максимов А.В., Оскорбин Н.М. Многопользовательские информационные системы: основы теории и методы исследования. — 2-е изд. испр. и доп. — Барнаул, 2013.
8. Носков С.И. Технология моделирования объектов с нестабильным функционированием и неопределенностью в данных. — Иркутск, 1996.
9. Поляк Б.Т., Назин С.А. Оценивание параметров в линейных многомерных системах с интервальной неопределенностью // Проблемы управления и информатики. — 2006. — № 1.
10. Zhilin S.I. Simple method for outlier detection in fitting experimental data under interval error // Chemometrics and Intellectual Laboratory Systems. — 2007. — Vol. 88. — No. 1.
11. Milanese M., Norton J., Piet-Lahanier H., Walter E. Bounding Approaches to System Identification (Eds). — New York, 1996.