УДК 621.513

Н.В. Кондрашова

СОГЛАСОВАНИЕ ВНЕШНЕГО КРИТЕРИЯ И СПОСОБА РАЗБИЕНИЯ ВЫБОРКИ ПРИ РЕШЕНИИ ЗАДАЧИ СТРУКТУРНО-ПАРАМЕТРИЧЕСКОЙ ИДЕНТИФИКАЦИИ МЕТОДОМ ГРУППОВОГО УЧЕТА АРГУМЕНТОВ

Введение

Детальному изучению проблемы разбиения выборки посвящено не очень много работ, что свидетельствует как о трудности ее анализа, так и о том, что при наличии некоторого опыта применения алгоритмов метода группового учета аргументов (МГУА) исследователи на практике обычно используют простые вычислительно малозатратные способы разбиения. При этом они учитывают специфику таких задач моделирования: идентификация, классификация, вычислительное доказательство или прогнозирование. Например, если необходимо аппроксимировать или классифицировать данные, то используют разбиение «по дисперсии»; если построить модель прогноза, применяют разбиение по хронологии; если протестировать модель или статистически обосновать скорость сходимости итерационного алгоритма МГУА, применяют случайное многократное разбиение, а результат моделирования усредняют. Есть резон в случайном разбиении также при построении статических моделей объекта, поскольку считается, что все уравнения связи в так называемой системе условных уравнений равновероятно могут быть отнесены как к обучающей A, так и к проверочной B выборке. Обычно исходную выборку W разбивают в следующем соотношении числа точек наблюдения: $n_A : n_B = 1:1$, если при этом вычисляется критерий несмещенности $\tilde{n}^2_{bias}$, поскольку обе выборки используются единообразно. Если в качестве внешнего критерия CR используется критерий регулярности $AR_{B|A}$, то отношение $n_A : n_B = 2:1$, эта эвристика связана с тем, что выборки имеют неодинаковое назначение. Вопрос о том, чем обосновано последнее соотношение, естественен. На наш взгляд, задача разбиения заслуживает углубленного изучения.

Задача о разбиении исходной выборки W на A и B по величине «дисперсий» [1] появилась после опубликования [2], которая способствовала ее обоснованию. В одной из поздних работ [3] А.Г. Ивахненко сформулировал двумерный аналог задачи разбиения строк матрицы $(\mathbf{X} \vdots \mathbf{y})$: задачу о выделении «ядра» — об оптимальном размере «скользящего окна» в виде квадратной матрицы входных переменных $\mathbf{X}$.

Далее рассмотрим только один вид разбиения выборки — разбиение по строкам наблюдений. Одно наблюдение эквивалентно строке значений переменных исходной матрицы, относящихся к одному моменту времени, одному паттерну или объекту.

В [4] предложено не ограничиваться разделением ранжированных точек наблюдений «по дисперсии», а выполнить полный перебор всех возможных составов множеств A и B (при условии $n_A = n_B$), для того чтобы найти множество моделей, в котором критерий минимума смещения параметров достигает наибольшего значения $\tilde{n}^2_{bias\max}$. Выписан алгоритм последовательной «переброски» точек из обучающей в проверочную выборку. После каждой «переброски» в процессе перебора моделей вычислялись приближенные значения параметров $\hat{\Theta}_f$ и критерия несмещенности. После этого выбиралась та модель $f(\mathbf{X}, \hat{\Theta}_f)$, для которой $\tilde{n}^2_{bias\max} \to \min$.

Работа [5] также была важна, так как содержала общую постановку задачи о наилучшем разбиении при использовании любого из множества применявшихся в то время внешних критериев. В ней лучшим считается разбиение, максимизирующее вероятность выбора «истинной» структуры модели, при которой достигается наименьшее значение функционала канонической формы внешнего критерия. Для реализации разбиения был предложен алгоритм построения поточечной функции распределения, обусловленной наличием шума, для конкретного внешнего критерия и каждого варианта разбиения.

Идеи, предложенные в двух последних работах, не получили широкого внедрения вследствие проблематичности их реализации на практике, так как в первой для каждого разбиения в полном переборе находится лучшая модель; во второй находится разбиение, максимизирующее вероятность получения «истинной» модели, из перебираемого множества вариантов разбиения. Кроме того, результат ощутимого увеличения точности и достоверности восстановления «истинной» структуры («истинного» сигнала) при значительном уровне шума не был получен.

В [6] предложено ρ^2-пропорциональное разбиение выборки, связанное с минимизацией среднеквадратической ошибки моделирования на всей выборке. В данной работе рассматривается также ρ-пропорциональное разбиение выборки. В связи с этим следует упомянуть работы [7, 8] о разбиении выборки в условиях моделирования при активном и повторном эксперименте.

1. Постановка задачи

И все же, при любом способе разбиений на A и B получаемая модель подстроена под разбиение, и для долгосрочного и среднесрочного прогнозирования реальных процессов она, как правило, не годится. Ее следует проверять на точках, не участвовавших ни в определении параметров, ни в определении структуры, т.е. на новой выборке C. В задаче моделирования возникает новый вопрос: как наилучшим образом разбить W на A, B и C ? Если точек мало, то, во-первых, можно, случайно разбив все множество W на $(A \cup B)$ и C, перебрать все возможные разбиения на A и B и найти модель

$$f_{s^*} = \arg\min_{f_s(\cdot)\in\Psi} CR(\mathbf{y}_Q, f_s(\mathbf{X}_Q, \hat{\Theta}_G)) \tag{1}$$

сложности s^*, где оценки параметров определяют как

$$\hat{\Theta}_G = \arg\min_{\Theta_G \in \Re^s} QR(\mathbf{y}_G, f_s(\mathbf{X}_G, \Theta_f)), \tag{2}$$

здесь Ψ — множество функций заданного класса структур моделей (полиномиальных, разностных и т.д.). Ограничимся линейной по параметрам моделью, в которой сложность s равна числу оцениваемых параметров, стоящих при различных аргументах, вычисленных с использованием $f_s(\cdot)$ — набора из s функций класса Ψ. $\Re^s$ — s-мерное множество действительных чисел. Если множество $Q = B$, то множество $G = A$ и, наоборот, если $Q = B$, то $G = A$; $Q \cap G = \varnothing$.

Во-вторых, можно многократно повторить случайное разбиение W на $(A \cup B)$ и C, а результат усреднить. Поскольку множество C получено случайным разбиением, то на этом можно закончить бесконечную цепочку последовательных разбиений. Так как в конечном итоге необходима единственная модель, то полученная и проверенная таким способом модель равновероятно имеет такие же характеристики точности на C, как и модели других случайных разбиений W на $(A \cup B)$ и C, где $(A \cup B) \cap C = \varnothing$.

В выражении (2) $QR(\cdot)$ — функционал качества решения задачи оценивания параметров, для которого при заданной структуре модели используется метод наименьших квадратов (МНК). Поскольку и на обучающей, и на проверочной подвыборках — это ошибка одной и той же модели, естественно потребовать, чтобы разбиение минимизировало ошибку, вычисленную на всей выборке. К этому требованию прилагаются условия, накладываемые на аддитивный шум выходного сигнала $\mathbf{y}$: он имеет нулевое математическое ожидание и конечную дисперсию σ^2. При поиске сложности структуры модели возможны три случая соотношений сложности модели (количества аргументов) s и числа наблюдений n: 1) $n < s$, 2) $n = s$ и 3) $n > s$.

Исходя из этих предпосылок, получено ρ^2-пропорциональное разбиение выборки для матриц $\mathbf{X}_A, \mathbf{X}_B$ полностолбцового ранга k=s:

$$\mathbf{X}_A^{\mathrm{T}}\mathbf{X}_A = \rho_B^2 \mathbf{X}_B^{\mathrm{T}}\mathbf{X}_B,\ \dim \mathbf{X}_A \neq \dim \mathbf{X}_B,\ n_A > s,\ n_B > s,\ n_A \neq n_B,\ 0 < s \leq m, \tag{3}$$

где s — число столбцов матриц (число переменных в усложняющихся моделях); $\rho_B \neq 0$ — некоторое число; размерность информационных матриц: $\dim \mathbf{X}_B^{\mathrm{T}}\mathbf{X}_B = \dim \mathbf{X}_A^{\mathrm{T}}\mathbf{X}_A = m \times m$; m — максимальное число аргументов (число столбцов).

Для матриц $\mathbf{X}_A, \mathbf{X}_B$ полнострочного ранга ($k = n$) с количеством строк $n < s$, чтобы найти их ρ^2-пропорциональность, запишем множество соотношений

$$\mathbf{X}_A \mathbf{X}_A^{\mathrm{T}} = \tilde{\rho}_B^2 \mathbf{X}_B \mathbf{X}_B^{\mathrm{T}},\ \dim \mathbf{X}_A = \dim \mathbf{X}_B,\ n = \min(n_A, n_B),\ n < s, n_A \neq n_B,\ 0 < s \leq m. \tag{4}$$

Если известна одна из подвыборок, например обучающая, спланировав эксперимент, можно создать проверочную выборку согласно соотношению

$$\mathbf{X}_A = \rho_B \mathbf{X}_B,\ \rho_B \neq 0,\ \dim \mathbf{X}_A = \dim \mathbf{X}_B,\ k = n_A = n_B = s\ \ 0 < s \leq m. \tag{5}$$

Вариант, когда одна из матриц $\mathbf{X}_A$, $\mathbf{X}_B$ полностолбцового, а другая полнострочного ранга, не рассматривается, так как при $k = n$ он сводится к (4). Случай повторного эксперимента ρ_B=1 является частным случаем (5).

Определение 1. ρ-оптимальными разбиениями выборки будем называть те разбиения, которые получаются в результате оптимизации функционалов, содержащих ρ-пропорциональные и ρ^2-пропорциональные зависимости данных подвыборок.

В задачах с эмпирическими данными соотношения (3)–(5), как правило, не выполняются. Поэтому поиск квазиоптимальных разбиений ℓ^*, μ^* или λ^* осуществляется минимизацией нормы разности в каждой из трех пар информационных матриц:

 ISSN 0572-2691

• для переопределенных матриц $\mathbf{X}_{A\ell}, \mathbf{X}_{B\ell}$

$$\ell^* = \min_{\rho\neq 0,\ \ell=\overline{1,\ L}} \| \mathbf{X}_{A\ell}^{\mathrm{T}}\mathbf{X}_{A\ell} - \rho_B^2 \mathbf{X}_{B\ell}^{\mathrm{T}}\mathbf{X}_{B\ell} \|, \ \dim \mathbf{X}_{A\ell} \neq \dim \mathbf{X}_{B\ell}, \ n_A \neq n_B > s; \tag{6}$$

• недоопределенных матриц

$$\mu^* = \min_{\rho\neq 0,\ \mu=\overline{1,\ L}} \| \mathbf{X}_{A\mu}\mathbf{X}_{A\mu}^{\mathrm{T}} - \tilde{\rho}_B^2 \mathbf{X}_{B\mu}\mathbf{X}_{B\mu}^{\mathrm{T}} \|, \ \dim \mathbf{X}_{A\mu} = \dim \mathbf{X}_{B\mu}, \ n_A \neq n_B < s; \tag{7}$$

• полностью определенных матриц

$$\lambda^* = \min_{\rho\neq 0,\ \lambda=\overline{1,\ \Lambda}} \| \mathbf{X}_{A\lambda} - \rho_B \mathbf{X}_{B\lambda} \|, \ \dim \mathbf{X}_{A\lambda} = \dim \mathbf{X}_{B\lambda}, \ n_A = n_B = s, \tag{8}$$

где $\|\bullet\|$ — некоторая матричная норма, не обязательно квадратичная.

Поскольку аффинные преобразования данных не изменяют результата оптимизации разбиения [9], то пропорциональной зависимости данных (6), (7) или (8), которая является разновидностью аффинного преобразования (гомотетией), переносом начала координат можно добиться того, что будет выполняться $\rho_B > 0$. Естественно, из условий минимума (6), (7) или (8) не следует, что при некоторой исходной матрице данных об объекте будет гарантировано найдена его «истинная» модель. Отметим, что при разбиении выборки в соответствии с (6) число n_A, а значит, и соотношение $n_A : n_B$, так же, как и наборы точек наблюдений, принадлежащие выборкам A и B, получаются автоматически, а при разбиении в соответствии с (7) и (8) — только наборы точек, поскольку в этом случае рассматриваются матрицы, у которых $k = s = n_A = n_B$, и, естественно, выполняется соотношение $n_A : n_B = 1:1$.

В [10] квадратичный критерий CR предложено представлять в компактной форме, как $CR(s) = \mathbf{y}^{\mathrm{T}}\mathbf{D}_{CR}(s)\mathbf{y}$, где $\mathbf{D}_{CR}(s)$ — симметрическая неотрицательно определенная матрица канонической формы квадратичного критерия CR, выражающаяся через единичную матрицу $\mathbf{I}$ и матрицу полностолбцового ранга $\mathbf{P}_{QG} = \mathbf{X}_Q(\mathbf{X}_G^{\mathrm{T}}\mathbf{X}_G)^{-1}\mathbf{X}_G^{\mathrm{T}}$. В качестве матриц с индексами Q, G могут быть матрица $\mathbf{X}_W$ и ее части $\mathbf{X}_A$ и $\mathbf{X}_B$, соответствующие выборкам A и B. В [5, 10, 11] получены матрицы канонических форм основных квадратичных критериев МГУА.

Известные алгоритмы МГУА, моделирующие объекты и процессы, не учитывают важность разбиения данных и способа формирования подвыборок при выборе критерия селекции. Обоснованность совместного применения способа разбиения данных и критерия внешнего дополнения при поиске наиболее точных моделей важна для правильного формирования правила останова в итерационных алгоритмах МГУА. Это обоснование в литературе о МГУА вообще не рассматривалось. В упомянутой работе [4] в численном эксперименте показано, что наилучшее разбиение выборки рекомендуется искать при максимизации критерия несмещенности. В данной работе этот внешний критерий всесторонне исследуется и теоретически обосновываются условия его применения при определенного вида ρ-оптимальных способах разбиения выборки.

2. ρ-оптимальное разбиение выборки для отбора моделей по минимуму критерия регулярности

В [6] доказано, что для получения оптимального разбиения выборки, при котором на подвыборках A и B обеспечивается неизменность структуры s и минимум критерия

$$\bar{J}_G(s) = \min_s \mathrm{M}[\| \mathbf{y}_G^0 - \hat{\mathbf{y}}_{GS} \|]^2, \ G = A, B, \ W = A \cup B \tag{9}$$

среднеквадратической ошибки отклонения выхода модели $\hat{\mathbf{y}}_G = \mathbf{X}_{Gs}\hat{\theta}_{As}$, усредненной по всем реализациям аддитивного шума в выходной переменной, необходимо создать или соблюдать условие ρ^2-пропорциональности информационных матриц. Обозначение $\mathrm{M}[\cdot]$ используется для символа математического ожидания; $\mathbf{y}_G^0$ — выход «истинной» модели; при этом выборкой G может быть любая из подвыборок A или B. Запись $\|\bullet\|^2$ используется для обозначения квадратичной нормы, например, $\|\mathbf{y}\|^2 = \mathbf{y}^{\mathrm{T}}\mathbf{y} = y_1^2 + ... + y_n^2$.

Поскольку среднее значение критерия регулярности $\overline{AR}(\cdot)$ является оценкой теоретической дисперсии ошибки модели $\bar{J}_B(s)$, то разбиение, удовлетворяющее ρ^2-пропорциональности данных, в пределе по множеству усреднений результатов численных экспериментов справедливо и для него, с точки зрения неизменности сложности структуры,

$$s^* = \arg\min \overline{AR}(\mathbf{X}_{As}, \mathbf{X}_B) = \arg\min \bar{J}_B(\mathbf{X}_{As}, \mathbf{X}_B),$$

где s^* — сложность оптимальной структуры при разбиении, удовлетворяющем (3). Предполагается некоррелированность различных реализаций шума между собой и с полезным сигналом, нулевое математическое ожидание и конечная дисперсия. Таким образом, при выборе моделей по минимуму критерия регулярности следует минимизировать критерии нормы квазиоптимального разбиения либо в силу большого перебора (6)–(8) применять менее эффективное «подобное по дисперсии» разбиение. Соотношение между квазиоптимальным и разбиением «по дисперсии» описано в [12].

При наиболее простом случае оценки параметров с помощью квадратных матриц $\mathbf{X}_A$ и $\mathbf{X}_B$ примем условие их одинаковой размерности, тогда аналогом условия (3) будет условие (5) ρ-пропорциональности. Размерность матриц $\mathbf{X}_A$ и $\mathbf{X}_B$ полного ранга $k = s$ равна $s \times s$ или, что то же, $n \times n$, где число столбцов $s = n_A = n_B = n$.

Исследуем критерий регулярности $AR_{B|A}(s)$ при условии ρ-пропорциональности данных и наличия аддитивного шума в выходной переменной

$$AR_{B|A} = \| \mathbf{y}_B - \mathbf{X}_B \hat{\Theta}_A \|^2 = \| \mathbf{y}_B^0 + \xi_B - \mathbf{X}_B(\mathbf{X}_A^{-1})(\mathbf{y}_A^0 + \xi_A) \|^2. \tag{10}$$

Операция поиска минимума по параметру s везде подразумевается, если не указывается другая. Здесь и далее опускаем ее и индекс s, чтобы упростить вывод формул. Сгруппируем регулярные и случайные компоненты. Усредним критерий по множеству реализаций шума и по множеству структур одинаковой сложности (равного числа аргументов и разных наборов), если шум не коррелирован с полезным сигналом:

$$\overline{AR}_{B|A} = \mathrm{M}[\| \mathbf{y}_B^0 - \mathbf{X}_B\mathbf{X}_A^{-1}\mathbf{y}_A^0 + (\xi_B - \mathbf{X}_B\mathbf{X}_A^{-1}\xi_A) \|^2] = \| \mathbf{y}_B^0 - \mathbf{X}_B\overline{\Theta}_A \|^2 +$$
$$+ \mathrm{M}[\| (\xi_B - \mathbf{X}_B(\mathbf{X}_A^{-1})\xi_A) \|^2] = AR_{B|A}^b + AR_{B|A}^v.$$

Очевидно, что структурная составляющая $AR_{B|A}^b$ является строго ниспадающей функцией при увеличении числа s. Рассмотрим «шумовую» составляющую

$$\begin{aligned} AR_{B|A}^v &= \mathrm{M}[(\xi_B - \mathbf{X}_B(\mathbf{X}_A^{-1})\xi_A)^{\mathrm{T}}(\xi_B - \mathbf{X}_B(\mathbf{X}_A^{-1})\xi_A)] = \mathrm{M}[\xi_B^{\mathrm{T}}\xi_B - \\ &- 2(\mathbf{X}_B\mathbf{X}_A^{-1}\xi_A)^{\mathrm{T}}\xi_B + \xi_A^{\mathrm{T}}(\mathbf{X}_B\mathbf{X}_A^{-1})^{\mathrm{T}}(\mathbf{X}_B\mathbf{X}_A^{-1})\xi_A] = \sigma_B^2 + \sigma_A^2 tr(\mathbf{I})/\rho_B^2. \end{aligned} \tag{11}$$

ISSN 0572-2691

При получении данного выражения использовано: условие некоррелированности различных реализаций шума между собой $\mathrm{cov}(\xi_A^{\mathrm{T}}\xi_B) = 0$; равенство $\mathbf{X}_B^{-1} = \rho_B \mathbf{X}_A^{-1}$ при подстановке в выражение $\xi_A^{\mathrm{T}}(\mathbf{X}_B\mathbf{X}_A^{-1})^{\mathrm{T}}(\mathbf{X}_B\mathbf{X}_A^{-1})\xi_A$, а также условие невырожденности матрицы $\mathbf{X}_A^{-1} (\det \mathbf{X}_A \neq 0)$.

Если дисперсии выборок A и B одинаковы, $\sigma^2 = \sigma_A^2 = \sigma_B^2$, то из (11) получим $AR_{A|B}^{v} = \sigma^2\left(1+\frac{1}{\rho_B^2}s\right) = \sigma^2\left(1+\frac{1}{\rho_B^2}n\right)$. Случайная составляющая является функцией, линейно возрастающей как от сложности структуры, так и от числа точек, поскольку $s=n=n_A=n_B$. Отсюда видно, что критерий регулярности является «адекватным» критерием.

Определение 2. «Адекватным» называется такой критерий, который с ростом дисперсии шума имеет глобальный минимум, сдвигающийся в сторону все более простых, не равных нулю структур.

Аналогичный результат может быть получен для математического ожидания критерия регулярности при ρ^2-пропорциональном разбиении выборки. Случайная составляющая при равенстве дисперсий на выборках A и B, а также для матрицы $\mathbf{X}_A$ полностолбцового ранга при использовании равенства $(\mathbf{X}_A^{\mathrm{T}}\mathbf{X}_A)^{-1}\rho_B^2 = (\mathbf{X}_B^{\mathrm{T}}\mathbf{X}_B)^{-1}$ имеет вид $AR_{A|B}^{v} = \sigma_B^2 + \sigma_A^2 \mathrm{tr}(\mathbf{P}_{BA}^{\mathrm{T}}\mathbf{P}_{BA})] = \sigma^2(1+\mathrm{tr}(\mathbf{P}_{AA})/\rho_B^2)$, где $\mathbf{P}_{AA} = \mathbf{X}_A(\mathbf{X}_A^{\mathrm{T}} \times \times \mathbf{X}_A)^{-1}\mathbf{X}_A^{\mathrm{T}}$, $\dim \mathbf{P}_{AA} = n_A \times n_A$.

Аналогично случайная составляющая для матрицы полнострочного ранга, если оценки параметров на A вычисляются через псевдообращение матрицы $\mathbf{X}_A$, имеет вид $AR_{A|B}^{v} = \sigma_B^2 + \sigma_A^2 \mathrm{tr}(\mathbf{H}_{BA}^{\mathrm{T}}\mathbf{H}_{BA})] = \sigma^2(1+\mathrm{tr}(\mathbf{H}_{BA}^{\mathrm{T}}\mathbf{H}_{BA}))$, где матрица $\mathbf{H}_{BA} = \mathbf{X}_B\mathbf{X}_A^{\mathrm{T}}(\mathbf{X}_A\mathbf{X}_A^{\mathrm{T}})^{-1}$.

Выводы об адекватности критерия справедливы и для ρ^2-пропорционального разбиения, а также в случае, близком к нему — квазиоптимального разбиения.

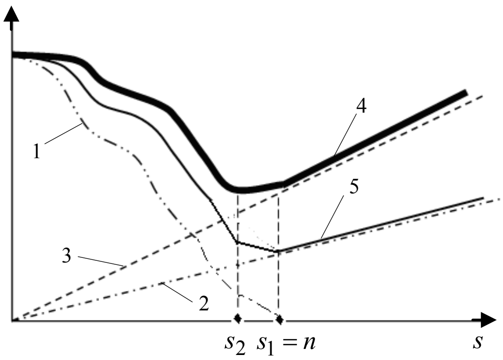

Рис. 1

На рис. 1 показано изменение критерия регулярности и его составляющих в зависимости от сложности структуры моделей при квазиоптимальном разбиении, где 1 — структурная составляющая; 2 — шумовая составляющая σ_1^2; 3 — шумовая составляющая σ_2^2; 4 — $AR(\sigma_2^2, s)$; 5 — $AR(\sigma_1^2, s)$, где $\sigma_2^2 > \sigma_1^2$.

Важным является то, что во всех случаях указанных разбиений и некоррелированном шуме при увеличении сложности структуры s или числа наблюдений n

достигается один минимум критерия регулярности. Квазиоптимальное разбиение есть результат минимизации нормы рассогласования ρ^2-пропорциональной зависимости информационных матриц. На рис. 1 показан сдвиг сложности структур, соответствующий минимуму критерия регулярности (сплошная жирная линия), в сторону более простых ($s_2 < s_1$) при увеличении дисперсии шума ($\sigma_2^2 > \sigma_1^2$), а также строго ниспадающий характер структурной составляющей и линейное возрастание при постоянной дисперсии шумовых составляющих в зависимости от числа s. Этот рисунок иллюстрирует «адекватность» критерия регулярности. Чем больше дисперсия шума, тем проще оптимальная модель. Структурная составляющая равна нулю при $s = n$. В данном случае для модели со сложностью структуры s_1, при $\sigma^2 = \sigma_1^2$, положение минимума критерия регулярности не изменилось по сравнению с минимумом его структурной составляющей. Здесь возможны два случая:

1) в матрице входных переменных **X**, кроме «истинных», присутствуют «неистинные» (случайные) аргументы, $m > s^0$;

2) в матрице **X** присутствуют только «истинные» аргументы, т.е. $m = s^0$.

Тогда в первом случае при уровне шума $\sigma^2 = \sigma_1^2$ модель со структурой s_1 является переусложненной, т.е. $s_1 > s^0$, так как случайный аргумент на входе имеет корреляцию с выходным сигналом, во втором найдена точная модель, а помеха на выходе отфильтрована (см. рис.1).

При дальнейшем увеличении дисперсии шума ($\sigma^2 = \sigma_2^2$) в первом случае модель со структурой s_2 ближе по сложности, чем со структурой s_1, к структуре «истинной» модели, если $s_2 > s^0$, и равна ей при $s_2 = s^0$. Во втором случае, когда присутствуют только «истинные» аргументы, метод и алгоритм, его реализующий, находят, что точнее будет более простая, чем со структурой s^0, адекватная шуму модель, у которой $s_2 < s^0$.

3. Влияние на адекватность критериев несмещенности ρ-оптимальных разбиений выборки

Любой усредненный критерий, для которого найдена матрица канонической формы, при диагональной матрице шума с дисперсией $\sigma^2 = \sigma_W^2 = \text{const}$ представляется в виде

$$\overline{CR}(s) = \mathrm{M}[CR(s)] = (\mathbf{y}^0)^{\mathrm{T}} \mathbf{D}_{CR}(s)\mathbf{y}^0 + \sigma^2 \mathrm{tr}\, \mathbf{D}_{CR}(s) = CR(s)^b + CR(s)^v,$$

где в случае матриц полностолбцового ранга $\mathbf{D}_{CR}(s) = \mathbf{I}_W - \mathbf{P}_{WW(CR)}$, а полнострочного ранга — $\mathbf{D}_{CR}(s) = \mathbf{I}_W - \mathbf{H}_{WW(CR)}$, $\dim \mathbf{D}_{CR}(s) = \mathbf{I}_W = n_W \times n_W$, $\mathbf{I}_W$ — единичная матрица; $CR(s)^b$ — структурная составляющая; $CR(s)^v$ — шумовая составляющая.

В данной работе в качестве примера подробно рассмотрен критерий, который нельзя представить в канонической форме, поскольку такая форма не найдена.

3.1. Отличительные особенности критериев несмещенности. Известно четыре разновидности критерия несмещенности.

Критерий несмещенности параметров (минимума смещения коэффициентов) [13] имеет вид

$$n_{bias(1)}^2(s) = \| \hat{\Theta}_{As} - \hat{\Theta}_{Bs} \|^2, \ \dim \hat{\Theta}_{As} = \dim \hat{\Theta}_{Bs} = s \times 1, \tag{12}$$

где МНК-оценки векторов параметров $\hat{\Theta}_{As}$ и $\hat{\Theta}_{Bs}$ определяются для каждой модели по своим выборкам в зависимости от выполнения следующих условий. Пусть k — ранг матрицы $\mathbf{X}_{Gs}$ размера $n_G \times s$.

Из [14] известно, что оценивание параметров осуществляется по-разному для матриц:

• полностолбцового ранга $k=s$, когда столбцы линейно независимы, при $n_G > s$ оценки имеют вид

$$\hat{\Theta}_{Gs} = (\mathbf{X}_{Gs}^{\mathrm{T}} \mathbf{X}_{Gs})^{-1} \mathbf{X}_{Gs}^{\mathrm{T}} \mathbf{y}_G; \;\; \dim(\mathbf{X}_G^{\mathrm{T}} \mathbf{X}_G) = s \times s, \;\; s = \overline{1, m}; \tag{13}$$

• полнострочного ранга $k = n_G$, когда строки линейно независимы и $n_G < s$, оценки получают через обращение псевдообратной матрицы, как

$$\hat{\Theta}_{Gs} = \mathbf{X}_{Gs}^{\mathrm{T}} (\mathbf{X}_{Gs} \mathbf{X}_{Gs}^{\mathrm{T}})^{-1} \mathbf{y}_G; \tag{14}$$

• квадратных, если и столбцы, и строки линейно независимы (что верно для невырожденных матриц), оценки параметров определяются как

$$\hat{\Theta}_{Gs} = (\mathbf{X}_{Gs})^{-1} \mathbf{y}_{Gs}, \; (n_{Gs} = s), \tag{15}$$

где выборкой G может быть A или B.

Другая более распространенная форма этого критерия — критерий несмещенности решений — имеет вид [15]

$$n_{bias(2)}^2(s) = \| \hat{\mathbf{y}}_W(\hat{\Theta}_{As}) - \hat{\mathbf{y}}_W(\hat{\Theta}_{Bs}) \|^2 = \| \mathbf{X}_{Ws}(\hat{\Theta}_{As} - \hat{\Theta}_{Bs}) \|^2. \tag{16}$$

При его вычислении ограничение на равенство сложности структур на выборках A и B сохраняется, нет ограничения на равенство объемов выборок, т.е. может быть $n_A \neq n_B$ и оценки $\hat{\mathbf{y}}_W(\cdot)$ вычисляются по данным X_W всей выборки W.

Критерии несмещенности (12) и (16) оперируют нормой отклонения оценок параметров на двух подвыборках A и B, $A \cup B = W$, $A \cap B = \varnothing$.

Абсолютно помехоустойчивый критерий [16] имеет вид

$$\begin{aligned} n_{bias(3)}^2(s) &= (\hat{\mathbf{y}}_W(\hat{\Theta}_{Ws}) - \hat{\mathbf{y}}_W(\hat{\Theta}_{As}))^{\mathrm{T}} (\hat{\mathbf{y}}_W(\hat{\Theta}_{Bs}) - \hat{\mathbf{y}}_W(\hat{\Theta}_{Ws})) = \\ &= (\hat{\Theta}_{Ws} - \hat{\Theta}_{As})^{\mathrm{T}} \mathbf{X}_{Ws}^{\mathrm{T}} \mathbf{X}_{Ws} (\hat{\Theta}_{Bs} - \hat{\Theta}_{Ws}). \end{aligned}$$

Его минимизация приводит к максимальной близости оценок коэффициентов модели, полученных на трех выборках: A, B и W. Название (не вполне удачное) связано с тем, что шумовая составляющая критерия линейно зависит от величины дисперсии σ^2, сложности модели s и не зависит от разбиения исходной матрицы только в случае σ^2 = const, в противном случае (если $\sigma_A^2 \neq \sigma_B^2$) можно показать, что критерий зависит от разбиения.

Критерий несмещенности ошибок [17], не являясь квадратичной нормой, имеет вид

$$n_{bias(4)}(s) = |AR_{W|A}(s) - AR_{W|B}(s)|. \tag{17}$$

Минимальные значения достигаются при их монотонном изменении с ростом сложности моделей s, когда под модулем меняется знак разности критериев регулярности, вычисленных на всей выборке по оценкам выборок A и B.

Для критериев несмещенности параметров $n^2_{bias(1)}(s)$ вида (12) и несмещенности ошибок $n_{bias(4)}(s)$ вида (17) не получены канонические формы, но любой внешний критерий, в том числе и усредненный критерий несмещенности, можно записать

$$\bar{n}^2_{bias(i)} = n^2_{bias(i)}{}^{b}(s) + n^2_{bias(i)}{}^{v}(s). \qquad (18)$$

Проанализируем усредненные формы критерия несмещенности параметров с точки зрения «адекватности» критерия при различных способах формирования подвыборок данных. Исследуем составляющие критерия (12) при условии, что будут выполняться:

а) ρ^2-пропорциональность; в) ρ-пропорциональность данных и лучшая модель будет получена при ρ-оптимальном разбиении выборки.

3.2. Критерий несмещенности параметров модели при ρ-оптимальном разбиении. При условии а) ρ^2-пропорциональности (3) и того, что $\mathbf{X}_A$ и $\mathbf{X}_B$ являются матрицами полностолбцового ранга, этот критерий имеет вид

$$n^2_{bias(1)}(s) = \| (\mathbf{X}^{\mathrm{T}}_{As}\mathbf{X}_{As})^{-1}\mathbf{X}^{\mathrm{T}}_{As}\mathbf{y}_A - (\mathbf{X}^{\mathrm{T}}_{Bs}\mathbf{X}_{Bs})^{-1}\mathbf{X}^{\mathrm{T}}_{Bs}\mathbf{y}_B \|^2 = \| (\mathbf{X}^{\mathrm{T}}_{As}\mathbf{X}_{As})^{-1} \times$$
$$\times \mathbf{X}^{\mathrm{T}}_{As}\mathbf{y}_A - \rho^2_B(\mathbf{X}^{\mathrm{T}}_{As}\mathbf{X}_{As})^{-1}\mathbf{X}^{\mathrm{T}}_{Bs}\mathbf{y}_B \|^2 = \| (\mathbf{X}^{\mathrm{T}}_{As}\mathbf{X}_{As})^{-1}(\mathbf{X}^{\mathrm{T}}_{As}\mathbf{y}_A - \rho^2_B\mathbf{X}^{\mathrm{T}}_{Bs}\mathbf{y}_B) \|^2. \qquad (19)$$

Если матрицы $\mathbf{X}_A$ и $\mathbf{X}_B$ являются матрицами полнострочного ранга, то критерий несмещенности параметров можно преобразовать к виду

$$n^2_{bias\ (1)}(s) = \| \mathbf{X}^{\mathrm{T}}_{As}(\mathbf{X}_{As}\mathbf{X}^{\mathrm{T}}_{As})^{-1}\mathbf{y}_A - \mathbf{X}^{\mathrm{T}}_{Bs}(\mathbf{X}_{Bs}\mathbf{X}^{\mathrm{T}}_{Bs})^{-1}\mathbf{y}_B \|^2 =$$
$$= \| \mathbf{X}^{\mathrm{T}}_{As}(\mathbf{X}_{As}\mathbf{X}^{\mathrm{T}}_{As})^{-1}\mathbf{y}_A - \tilde{\rho}^2_B\mathbf{X}^{\mathrm{T}}_{Bs}(\mathbf{X}_{As}\mathbf{X}^{\mathrm{T}}_{As})^{-1}\mathbf{y}_B \|^2;$$

При условии б) ρ-пропорциональности данных рассмотрим случай, когда выполняется условие $\mathbf{X}_A = \rho_B\,\mathbf{X}_B$, т.е. $s = n_A = n_B$. Критерий несмещенности (19) при условии, что $(\mathbf{X}_{Bs})^{-1} = \rho_B(\mathbf{X}_{As})^{-1}$, перепишем в виде

$$n^2_{bias(1)}(s) = \| (\mathbf{X}_{As})^{-1}\mathbf{y}_A - (\mathbf{X}_{Bs})^{-1}\mathbf{y}_B) \|^2 = \| (\mathbf{X}_{As})^{-1}(\mathbf{y}_A - \rho_B\mathbf{y}_B) \|^2. \qquad (20)$$

Найдем математическое ожидание (20) при условии некоррелированности сигнала с шумом и усреднении по множеству реализаций шума, если $\mathbf{y}_W = \mathbf{y}^0_W + \xi_W$. В результате получим

$$\bar{n}^2_{bias(1)}(s) = \mathrm{M}\| (\mathbf{X}_{As})^{-1}[(\mathbf{y}^0_A - \rho_B\mathbf{y}^0_B) + (\xi_A - \rho_B\xi_B)] \|^2 = n^2_{bias(1)}(s)^b + n^2_{bias(1)}(s)^v.$$

Если матрица $\mathbf{X}_{As}$ содержит только «истинные» аргументы, то при $m = s = s^0$, ввиду того, что $\Theta^0_{As} = \Theta^0_{Bs} = \Theta^0$, будет обеспечено равенство нулю структурной составляющей, т.е. $\| \overline{\Theta}_{As^0} - \overline{\Theta}_{Bs^0} \|^2 = 0$, так как будут выполнены равенства $\overset{o}{\mathbf{y}}_A = \overset{0}{\mathbf{X}}_A\,\Theta^0$, $\overset{o}{\mathbf{y}}_B = \overset{0}{\mathbf{X}}_B\,\Theta^0$, где Θ^0 — вектор параметров «истинной» модели. При неполном составе «истинных» аргументов на входе ($m < s^0$) выполняется $\overline{\Theta}_{As} \cong \overline{\Theta}_{Bs} \neq \Theta^0$, $\| \overline{\Theta}_{As} - \overline{\Theta}_{Bs} \|^2 \cong 0$. Если $\sigma^2_A \neq \sigma^2_B$, то шумовая составляющая равна

$$n^2_{bias\ (1)}(s)^v = (\sigma^2_A + \rho^2_B\sigma^2_B)\,\mathrm{tr}\,[(\mathbf{X}^{-1}_{As})^{\mathrm{T}}(\mathbf{X}^{-1}_{As})] = (\sigma^2_A + \rho^2_B\sigma^2_B)\,\mathrm{tr}\,(\mathbf{X}_{As}\mathbf{X}^{\mathrm{T}}_{As})^{-1}.$$

 ISSN 0572-2691

Она зависит от суммы дисперсий в подвыборках A и B, и если дисперсии равны ($\sigma_A^2 = \sigma_B^2 = \sigma^2$), то σ^2 имеет коэффициент $(1+\rho_B^2)$, который для конкретной выборки $1 < (1+\rho_B^2) = \text{const}$. След матрицы $\operatorname{tr}[(\mathbf{X}_{As}^{-1})^{\mathrm{T}} \mathbf{X}_{As}^{-1}] = \operatorname{tr}(\mathbf{X}_{As}\mathbf{X}_{As}^{\mathrm{T}})^{-1} =$ $= \operatorname{tr}(\mathbf{X}_{Bs}\mathbf{X}_{Bs}^{\mathrm{T}})^{-1} / \tilde{\rho}_B^2$ при выполнении (4) и увеличении размера матрицы $s \times s$ или, что то же, $n \times n$ растет. Шумовая составляющая для пропорциональных подвыборок с ростом сложности s и числа наблюдений n растет. Таким образом, критерий несмещенности параметров при минимизации критерия разбиения и ρ-пропорциональности данных подвыборок не является «адекватным» критерием, поскольку, при его минимизации (квазиоптимизации) выбирается тривиальная модель минимальной (нулевой) структуры.

Вернемся к а) ρ^2-пропорциональности данных. Проанализируем (19) для матриц полностолбцового ранга

$$n_{bias\,(1)}^2(s) = \| (\mathbf{X}_{As}^{\mathrm{T}}\mathbf{X}_{As})^{-1}[\mathbf{X}_{As}^{\mathrm{T}}(\mathbf{y}_A^0 + \xi_A) - \rho_B^2 \mathbf{X}_{Bs}^{\mathrm{T}}(\mathbf{y}_B^0 + \xi_B)] \|^2.$$

Усредняя помеху по множеству реализаций шума и множеству моделей, имеющих одинаковую сложность структуры, получаем

$$\bar{n}_{bias\,(1)}^2(s) = \mathrm{M}\, \| (\mathbf{X}_{As}^{\mathrm{T}}\mathbf{X}_{As})^{-1}[(\mathbf{X}_{As}^{\mathrm{T}}\, \mathbf{y}_A^0 - \rho_B^2 \mathbf{X}_{Bs}^{\mathrm{T}}\, \mathbf{y}_B^0)] \|^2 + \mathrm{M}\, \| (\mathbf{X}_{As}^{\mathrm{T}}\mathbf{X}_{As})^{-1} \times$$

$$\times [(\mathbf{X}_{As}^{\mathrm{T}}\xi_A - \rho_B^2 \mathbf{X}_{Bs}^{\mathrm{T}}\xi_B)] \|^2 = \| \bar{\Theta}_{As} - \bar{\Theta}_{Bs} \|^2 + \mathrm{M}\,\{[(\mathbf{X}_{As}^{\mathrm{T}}\, \mathbf{X}_{As})^{-1}\mathbf{X}_{As}^{\mathrm{T}}\, \xi_A]^{\mathrm{T}} \times$$

$$\times (\mathbf{X}_{As}^{\mathrm{T}}\mathbf{X}_{As})^{-1}\mathbf{X}_{As}^{\mathrm{T}}\, \xi_A\} + \mathrm{M}\,\{[\rho_B^2(\mathbf{X}_{As}^{\mathrm{T}}\mathbf{X}_{As})^{-1}\mathbf{X}_{Bs}^{\mathrm{T}}\, \xi_B]^{\mathrm{T}} \rho_B^2 (\mathbf{X}_{As}^{\mathrm{T}}\mathbf{X}_{As})^{-1}\mathbf{X}_{Bs}^{\mathrm{T}}\, \xi_B\} =$$

$$= \| \bar{\Theta}_{As} - \bar{\Theta}_{Bs} \|^2 + \sigma_A^2 \operatorname{tr}(\mathbf{V}_{AA}^{\mathrm{T}}\, \mathbf{V}_{AA}) + \rho_B^4 \sigma_B^2 \operatorname{tr}(\mathbf{V}_{AB}^{\mathrm{T}}\, \mathbf{V}_{AB}) = \bar{n}_{bias\,(1)}^2(s)^b + \bar{n}_{bias\,(1)}^2(s)^v,$$

$$\mathbf{V}_{AA} = (\mathbf{X}_{As}^{\mathrm{T}}\mathbf{X}_{As})^{-1}\mathbf{X}_{As}^{\mathrm{T}}, \dim \mathbf{V}_{AA} = s \times n_A, \mathbf{V}_{AB} = (\mathbf{X}_{As}^{\mathrm{T}}\mathbf{X}_{As})^{-1}\mathbf{X}_{Bs}^{\mathrm{T}}, \dim \mathbf{V}_{AB} = s \times n_B.$$

Если дисперсия на выборках A и B одинакова $\sigma^2 = \sigma_A^2 = \sigma_B^2$, то шумовая составляющая имеет вид

$$\bar{n}_{bias\,(1)}^2(s)^v = \sigma^2[\operatorname{tr}(\mathbf{V}_{AA}^{\mathrm{T}}\mathbf{V}_{AA}) + \rho_B^4 \operatorname{tr}(\mathbf{V}_{AB}^{\mathrm{T}}\mathbf{V}_{AB})].$$

При выводе использовалось свойство некоррелированности различных реализаций шума между собой и с полезным сигналом. Усредненная структурная составляющая $\bar{n}_{bias\,(1)}^2(s)^b$ имеет такой же вид $\| \bar{\Theta}_{As} - \bar{\Theta}_{Bs} \|^2$, как и в случае ρ-пропорциональности данных, но значения параметров, вычисленные с помощью формул (13) или (14), отличаются от вычисленных по квадратным матрицам (15). Чтобы разницы коэффициентов на различных выборках при одних и тех же переменных, и одном и том же значении s не были равны нулю, необходимо, чтобы выборки были максимально неподобны между собой. Поэтому если применяется критерий несмещенности, то необходимо строить модели по данным максимально различающихся (неподобных) выборок. Например, в качестве критерия разбиения использовать не минимум нормы (3), а ее максимум:

$$\ell^* = \arg \max_{\rho_{B\ell}^2 \neq 0,\ \ell = \overline{1, L}} \| \mathbf{X}_{A\ell}^{\mathrm{T}}\mathbf{X}_{A\ell} - \rho_{B\ell}^2 \mathbf{X}_{B\ell}^{\mathrm{T}}\mathbf{X}_{B\ell} \|, \tag{21}$$

где ℓ^* — наилучшее разбиение, либо для получения разбиения лучшего l^* использовать «неподобные по дисперсии» подвыборки A и B:

$$l^* = \arg\max_{\overline{l=1,\,L_d}} \operatorname{tr}(\widetilde{\mathbf{X}}_{Al}^{\mathrm{T}}\widetilde{\mathbf{X}}_{Al} - \widetilde{\mathbf{X}}_{Bl}^{\mathrm{T}}\widetilde{\mathbf{X}}_{Bl}) = \arg\max_{\overline{l=1,\,L_d}} [\operatorname{tr}(\widetilde{\mathbf{X}}_{Al}^{\mathrm{T}}\widetilde{\mathbf{X}}_{Al}) - \operatorname{tr}(\widetilde{\mathbf{X}}_{Bl}^{\mathrm{T}}\widetilde{\mathbf{X}}_{Bl})], \quad (22)$$

где для вычисления «дисперсий» используются диагональные элементы информационных матриц, элементы которых центрированы как $\widetilde{x}_{jj} = x_{jj} - \overline{x}_j$, $j = \overline{1, m}$, а $\overline{x}_j$ — средние значения аргументов, $j = \overline{1, m}$, вычисляются по данным столбцов матриц $\mathbf{X}_{Al}$ и $\mathbf{X}_{Bl}$. Индекс l обозначает набор строк при переборе содержания матриц $\mathbf{X}_{Al}$ и того, что $\mathbf{X}_{Wl}^T = (\mathbf{X}_{Al}^T \vdots \mathbf{X}_{Bl}^T)$, $A \cap B = \varnothing$.

Когда присутствует все множество «истинных» аргументов и матрицы $\mathbf{X}_{As}$ и $\mathbf{X}_{Bs}$ являются матрицами полного ранга, то при достижении $s = s^0$ и $n_A \geq s$, $n_B \geq s$ норма ошибок векторов параметров $\|\hat{\Theta}_{As} - \hat{\Theta}_{Bs}\|^2 = \|\Theta_A^0 - \Theta_B^0\|^2 = 0$ (так как $\hat{\Theta}_{As} = \hat{\Theta}_{Bs} = \Theta^0$). При $s \neq s^0$, т.е. при неполном составе «истинных» аргументов на входе ($m < s^0$), выполняется $\overline{\Theta}_{As} \cong \overline{\Theta}_{Bs} \neq \Theta^0$. С ростом числа «истинных» аргументов в модели немонотонно изменяющаяся норма приближается к нулю: $\|\overline{\Theta}_{As} - \overline{\Theta}_{Bs}\|^2 \cong 0$. Шумовая составляющая матриц полнострочного ранга имеет вид

$$\overline{n}_{bias\,(1)}^2(s)^v = \sigma_A^2 \operatorname{tr}(\mathbf{U}_{AA}^{\mathrm{T}}\mathbf{U}_{AA}) + \sigma_B^2 \operatorname{tr}(\mathbf{U}_{BB}^{\mathrm{T}}\mathbf{U}_{BB}) = \sigma_A^2 \operatorname{tr}(\mathbf{X}_{As}\mathbf{X}_{As}^{\mathrm{T}})^{-1} + \sigma_B^2 \operatorname{tr}(\mathbf{X}_{Bs}\mathbf{X}_{Bs}^{\mathrm{T}})^{-1},$$

а при равенстве дисперсий на A и B $\overline{n}_{bias\,(1)}^2(s)^v = \sigma^2[\operatorname{tr}(\mathbf{X}_{As}\mathbf{X}_{As}^{\mathrm{T}})^{-1} + \operatorname{tr}(\mathbf{X}_{Bs}\mathbf{X}_{Bs}^{\mathrm{T}})^{-1}]$ монотонно растет с ростом сложности структуры модели. Здесь $\mathbf{U}_{GG} = \mathbf{X}_{Gs}^{\mathrm{T}}(\mathbf{X}_{Gs}\mathbf{X}_{Gs}^{\mathrm{T}})^{-1}$, $\dim \mathbf{U}_{GG} = s \times n_G$, где матрица G может быть A или B. С учетом (4), из которого вытекает $\widetilde{\rho}_B^2(\mathbf{X}_{As}\mathbf{X}_{As}^{\mathrm{T}})^{-1} = (\mathbf{X}_{Bs}\mathbf{X}_{Bs}^{\mathrm{T}})^{-1}$, получаем $\overline{n}_{bias\,(1)}^2(s)^v = \sigma^2(1 + \rho_B^2)\operatorname{tr}(\mathbf{X}_{As}\mathbf{X}_{As}^{\mathrm{T}})^{-1}$.

При ρ^2-пропорциональности данных, матрицах $\mathbf{X}_{As}$ и $\mathbf{X}_{Bs}$ полного ранга ($s \leq n_A$, $s \leq n_B$) минимум критерия несмещенности параметров с увеличением дисперсии шума соответствует тривиальной модели при всех рассмотренных соотношениях размеров матриц. При максимально неподобных выборках и увеличении шума в данных геометрическое место минимумов критерия сдвигается в сторону более простых моделей $s^* \leq s^0$, однако минимумы с увеличением s и ростом σ^2 могут перемещаться в разных направлениях (рис. 2 и рис. 3). Таким образом, в целом, помимо влияния уровня и степени корреляции шума в данных, критерий несмещенности параметров, если выполнены условия пропорциональности данных, не является «адекватным» критерием, а является таковым при максимизации ρ-оптимального разбиения (21) или (22).

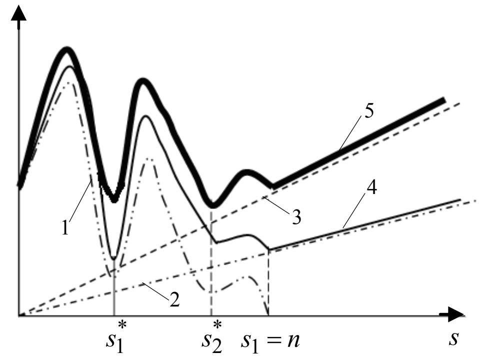

Рис. 2

На рис. 2 показано изменение критерия несмещенности параметров и его составляющих в зависимости от сложности структуры моделей и дисперсии шума при ρ-оптимальном разбиении данных. Сдвиг сложности структур, соответствующий минимуму критерия несмещенности (сплошные линии), в сторону более простых структур ($s_1^* < s_2^* < s_1$) при увеличении дисперсии шума, где 1 — структурная составляющая; 2 — шумовая составляющая σ_1^2; 3 — шумовая составляющая σ_2^2; 4 — $n_{bias(1)}^2(\sigma_1^2, s)$; 5 —$n_{bias(1)}^2(\sigma_2^2, s)$, $\sigma_2^2 > \sigma_1^2$. Показано линейное возрастание шумовых составляющих в зависимости от числа s при постоянной дисперсии. Усредненная структурная составляющая равна нулю при отсутствии шума и близка к нулю при одинаковом уровне шума во всех рассмотренных случаях соотношений размеров матриц, по данным которых оцениваются параметры модели.

В иллюстрируемом примере минимум критерия несмещенности для модели со сложностью структуры s_1^*, имеющей шум в данных с дисперсией $\sigma_1^2 < \sigma_2^2$, расположен левее положения s_2^*, соответствующего большей дисперсии шума. Положение минимумов может поменяться при увеличении σ^2 из-за произвольного характера изменения структурной составляющей при неодновременном скачке сложности на подвыборках с разной дисперсией, причем оптимальная структура, отвечающая меньшему уровню шума, может иметь большее значение критерия и быть более сложной и, наоборот. В отличие от качественного характера зависимостей, представленных на рис. 2, рис. 3, иллюстрирует результаты численного эксперимента. Здесь показано изменение минимума критерия с увеличением уровня шума в обеих частях выборки при поиске оптимальной структуры, если существует разница уровня шума на подвыборках A и B. Показаны такие зависимости при разнице уровней шума 10 % и 20 %. Чем больше эта разница, тем более плавно изменяется зависимость значений критерия от уровня шума. При нулевом уровне шума в выборке $\boldsymbol{A}$ и превышении шума на 10 % в выборке B выбирается «истинная» структура модели (s =11), а при превышении шума на 20 % тоже выбирается «истинная» структура, но значения параметров при этом имеют значительное отклонение от «истинных», так как величина критерия возросла приблизительно в 10 раз. Если шум в данных на подвыборке A составляет 30 % от диапазона изменения x_j, а на B — 40 % (разница между ними также 10 %), то минимум критерия достигается для моделей сложности (s = 10) при увеличении критерия более, чем на два порядка и т.д. Заметим, что при этом не выбирается тривиальная модель с нулевой структурой.

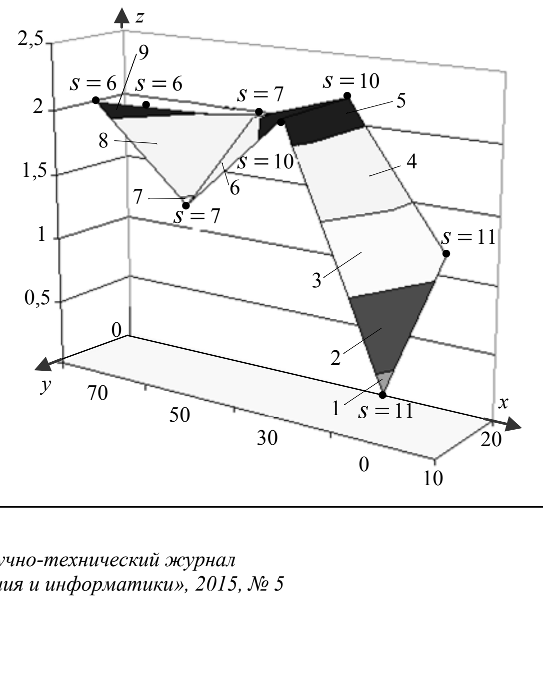

Рис. 3

На рис. 3 показано изменение минимума критерия несмещенности параметров, соответствующего оптимальной модели, при увеличении уровня шума в A и B для двух значений разницы уровней между выборками A и B (двух мер «неподобия по дисперсии»), где ось $0x$ — уровень шума, %; ось $0y$ — разница уровней шума на A и B, %; ось $0z$ — $\lg(n^2_{bias(1)})$. Пределы изменения этого логарифма: 1 — область 0–0,5; 2 — 0,5–1; 3 = 7 — 1–1,5; 6 = 8 = 4 — 1,5–2; 5 = 9 — 2–2,5.

Заключение

Доказано, что критерий регулярности адекватен шуму при ρ-пропорциональности данных, квазиоптимальности их разбиения или «подобном по дисперсии разбиении».

Доказано и подтверждено в численном эксперименте, что критерий несмещенности параметров при квазиоптимальном разбиении в общем случае не является «адекватным» шуму, так как выбирает тривиальную модель, что необходимо учитывать при формировании правила останова в алгоритмах МГУА.

При выборе моделей по минимуму критерия несмещенности параметров следует избегать квадратичной и линейной пропорциональности данных, квазиоптимального разбиения и производить «подобное по дисперсии» разбиение, а применять, например, максимум разности ρ-оптимального разбиения информационных матриц либо «неподобное по дисперсии» разбиение.

Н.В. Кондрашова

УЗГОДЖЕННЯ ЗОВНІШНЬОГО КРИТЕРІЮ І СПОСОБУ РОЗБИТТЯ ВИБІРКИ ПРИ РОЗВ'ЯЗАННІ ЗАДАЧІ СТРУКТУРНО-ПАРАМЕТРИЧНОЇ ІДЕНТИФІКАЦІЇ МЕТОДОМ ГРУППОВОГО УРАХУВАННЯ АРГУМЕНТІВ

Досліджується обґрунтованість спільного застосування способу розбиття даних і критерію зовнішнього доповнення при пошуці найбільш точних моделей. Описано розбиття для всіх співвідношень розмірів вихідних матриць. Досліджуються зовнішні критерії МГУА, теоретично обґрунтовується їх застосування для певного виду оптимальних способів розбиття вибірки. Теоретично доказано і підтверджено у чисельному експерименті, що критерій незміщеності параметрів, якщо виконані умови пропорційності даних, не є «адекватним» шуму при мінімізації критерію розбиття, а лише при його максимізації.

N.V. Kondrashova

MATCHING OF EXTERNAL CRITERION AND METHOD OF SAMPLE PARTITIONING FOR SOLVING PROBLEM OF STRUCTURAL-PARAMETRIC IDENTIFICATION BY GROUP METHOD OF DATA HANDLING

We investigate the validity of the joint application of the data partitioning method and criterion of external additions in finding the most accurate models. Quasi-optimal partitions for all ratios of sizes of the initial matrices are described. GMDH external criteria are comprehensively investigated. Their use in certain types of optimal ways of partitioning the sample is theoretically justified. If the condition of proportionality of data is satisfied, the theoretical substantiation and confirmation in numerical experiments were obtained that the criterion of parameters unbiased ness, is not "adequate" to noise while minimizing the criterion of the sample division, it takes place only when it is maximized.

1. *Ивахненко А.Г.* Системы эвристической самоорганизации в технической кибернетике. — Киев: Техніка, 1971. — 372 с.
2. *Павлов О.А.* Критерій ранжування для порогового самовідбору змінних в алгоритмах МГУА // Автоматика.. — 1969. — № 4. — С. 89–91.
3. *Ivakhnenko A.G., Savchenko E.A., Ivakhnenko G.A.* Discrete optimization of square data samples as the first stage of the optimization of a discriminant or predictive model // Pattern Recognition and Image Analysis. — 2004. — **14.** — P. 489–494.
4. *Висоцький В.М.* Про найкращий поділ вихідних даних в алгоритмах МГУА // Автоматика. — 1976. — № 3. — С. 71–74.
5. *Юрачковский Ю.П., Грошков А.Н.* Оптимальное разбиение исходных данных на обучающую и проверочную последовательности на основе анализа функции распределения критерия // Там же. — 1980. — № 2. — С. 5–9.
6. *Степашко В.С.* Структурная идентификация прогнозирующих моделей в условиях планируемого эксперимента // Там же. — 1992. — № 1. — С. 26–35.
7. *Сарычев А.П.* Решение проблемы разбиения в МГУА при расчете критерия регулярности в условиях активного эксперимента // Там же. — 1989. — № 4. — С. 19–27.
8. *Сарычев А.П.* Определение J-оптимального множества регрессоров по повторным выборкам наблюдений // Там же. — 1993. — № 3. — С. 58–66.
9. *Степашко В.С., Кондрашова Н.В.* Оценивание трансформации геометрических фигур. — Праці Міжнародного семінару з індуктивного моделювання (МСІМ–2005). Київ, 2005. — Київ: МННЦ ІТС, 2005. — С. 294–301.
10. *Юрачковский Ю.П., Грошков А.Н.* Применение канонических форм внешних критериев для исследования их свойств // Автоматика. — 1979. — № 3. — С. 85–89
11. *Ивахненко А.Г., Юрачковский Ю.П.* Моделирование сложных систем по экспериментальным данным. — М.: Радио и связь, 1987. — 120 с.
12. *Степашко В.С., Кондрашова Н.В.* Анализ проблемы разбиения выборки для алгоритмов МГУА // Кибернетика и вычисл. техника. — 2002. — Вып. 136. — С. 3–15.
13. *Ивахненко А.Г.* Долгосрочное прогнозирование и управление сложными системами — Киев: Техніка, 1975. — 312 с.
14. *Гантмахер Ф.Р.* Теория матриц: 3-е изд. — М.: Наука, 1967. — 575 р.
15. *Светальский Б.К.* Алгоритм и программа моделирования сложных систем по критерию несмещенности и баланса переменных для системного многократного дифференциального прогноза и управления // Прямой синтез математических моделей сложных объектов при небольшом числе экспериментальных данных. — Киев: Наук. думка, 1974. — С. 42–44.
16. *Ивахненко А.Г., Высоцкий В.Н., Ивахненко Н.А.* Основные разновидности критерия минимума смещения модели и исследование их помехоустойчивости // Автоматика. — 1978. — № 1. — С. 32–53.
17. *Ivakhnenko A.G., Ivakhnenko G.A., Savchenko E.A.* GMDH algorithm for optimal model choice by the external error criterion with the extension of definition by model bias and its applications to the committees and Neural Networks // Pattern Recognition and Image Analysis. — 2002. — **12**, N 4. — P. 347–353.

Получено 10.12.2014
После доработки 26.03.2015