УДК 681.513.8

Н.В. Кондрашова

Влияние способа разбиения выборки в алгоритмах метода группового учета аргументов на адекватность критерия внешнего дополнения

Рассмотрены способы разбиения выборки исходных данных в алгоритмах метода группового учета аргументов. Проведен анализ соответствия квазиоптимального метода разбиения подвыборок и критерия внешнего дополнения при выборе наилучшей структуры модели. Численный эксперимент показывает, что такие разбиения способствуют поиску «истинных» аргументов моделей в алгоритмах указанного метода с минимизацией критерия несмещенности решений.

Розглянуто способи розбиття вибірки вихідних даних в алгоритмах методу групового обліку аргументів. Проведено аналіз відповідності квазіоптимального методу розбиття підвибірок і критерію зовнішнього доповнення при виборі найкращої структури моделі. Чисельний експеримент показує, що такі розбиття сприяють пошуку «істинних» аргументів моделей в алгоритмах вказаного методу з мінімізацією критерію незміщеності рішень.

Введение. Известные алгоритмы МГУА для моделирования объектов и процессов не учитывают важность разбиения данных и способа формирования подвыборок при выборе критерия селекции. Вопрос обоснования совместного применения способа разбиения данных и критерия внешнего дополнения при поиске моделей в литературе широко не рассматривался. Помимо случайного разбиения, широко применяемым, есть также способ разбиения по величине *дисперсии* выборки [1]. В [2] исследован алгоритм поиска наилучшего разбиения выборки при использовании в качестве внешнего – критерия несмещенности. Для этого в процессе перемешивания точек в подвыборках, при условии равенства их длин, отбираются для каждого разбиения модели с наибольшим значением критерия несмещенности параметров, а из них выбирается модель с наименьшим его значением. В [3] максимизируется вероятность получения *истинной* модели при переборе множества вариантов разбиения. Работа [4] в качестве наилучшего разбиения при минимизации критерия регулярности предлагает: в теории – квадратичное (ρ^2-пропорциональное), а на практике – квазиоптимальное разбиение выборки. В [5] исследуются свойства наиболее распространенных критериев внешнего дополнения при квазиоптимальном разбиении. В статье рассмотрено:

- использование критерия минимума несмещенности решений в качестве внешнего дополнения и различных способов разбиения;
- какой вид должен иметь критерий разбиения при использовании того или иного способа разбиения исходной выборки.

Исследуем взаимосвязь и отличительные особенности известных способов разбиения выборки.

Разбиение по дисперсии

Пусть дана исходная выборка W в виде матрицы $(\mathbf{X}\vdots\mathbf{y})$. Для вычисления этого разбиения ранжируют по величине дисперсии строки матрицы входных переменных $\mathbf{X}$ размерности $n_W \times m$, где n_W – число строк (наблюдений), m – число столбцов (входных переменных). Сначала определяются выборочные средние в столбцах $\overline{x_i} = \frac{1}{n_W}\sum_{j=1}^{n_W} x_{ji}$, $i = \overline{1,m}$, т.е. оценку математического ожидания каждой переменной. После этого вычисляются величины сумм квадратов отклонений от этих средних (дисперсии) в каждой точке наблюдения:

$$\eta_j^2 = \sum_{i=1}^{m}(x_{ji} - \overline{x_i})^2, \quad j = \overline{1, n_W},$$

затем все полученные значения η_j^2 ранжируются в порядке убывания. Деление η_j^2 на $(m-1)$ не, проводится, как при расчете выборочной

дисперсии, вследствие применения к членам ряда отношений « > », « < ». Кроме того, в отличие от выборочной дисперсии здесь суммируются по строке квадраты отклонений каждого элемента матрицы от средних значений своего столбца. При вычислении такой дисперсии происходит абстрагирование от смыслового содержания переменных и учитывается только то, насколько в каждом конкретном наблюдении, зафиксированном в исходной выборке, входные переменные отклонились от своих средних значений. Для того чтобы выполнялось условие сбалансированности подвыборок по величине, наиболее часто применяемыми отношениями размеров подвыборок $n_A:n_B$ считаются:

- 1:1, которое связано с подбором подвыборок A и B равного объема, при проектировании точек на прямую линию;
- 2:1, обеспечивающее поиск срединного (медианного) расстояния между центрами множеств S_1, S_2 и S_3 при проектировании их на плоскость в виде трех равномощных множеств, где $A=S_i\cup S_j$, $B=S_k$, $i,j,k\in\{1,2,3\}$, $i\neq j\neq k$.

Поэтому такие разбиения сбалансированы с точностью до геометрии представления множеств разбиения. В [1] точки, ранжированные по убыванию значений дисперсии, относят в подвыборки A и B *подобных* или *неподобных по дисперсии* разбиений двумя способами:

- через одну или две точки, что означает получение частей выборки, разной степени подобия, если заданы отношения $n_A:n_B$, равные 1:1 либо 2:1;
- первые $\lfloor 2/3\, n_W \rfloor$ точек с бóльшей дисперсией – в обучающую подвыборку, оставшаяся $\lfloor 1/3\, n_W \rfloor$ часть точек – в проверочную, что приводит к получению неподобных (непохожих) по дисперсии подвыборок, где $\lfloor\cdot\rfloor$ означает операцию округления.

Квадратичное или ρ^2-пропорциональное разбиение выборки получают в соответствии с отношением [4]

$$\mathbf{X}_A^T\mathbf{X}_A=\rho_B^2\mathbf{X}_B^T\mathbf{X}_B,\ \ \rho_B^2\neq 0. \qquad (1)$$

Обозначим информационные матрицы, как $\chi_A=\mathbf{X}_A^T\mathbf{X}_A$ и $\chi_B=\mathbf{X}_B^T\mathbf{X}_B$. Разбиение по дисперсии в терминах ρ^2-пропорциональной зависимости подвыборок можно интерпретировать так.

Пусть выборка такова, что все $\bar{x}_i=0, i=\overline{1,m}$, тогда $\tilde{\eta}_j^2\overset{\Delta}{=}\eta_j^2\,|_{\bar{x}_i=0}=\sum_{j=1}^{m}(x_{ji})^2$, где x_{ji} – центрированные переменные. В этом случае модель не содержит свободного члена, т.е. $\theta_0=0$. Далее легко получить, что

$$tr(\mathbf{X}^T\mathbf{X})=\sum_{i=1}^{m}\chi_{ii}=\sum_{i=1}^{m}\sum_{j=1}^{n_W}(x_{ji})^2=\sum_{j=1}^{n_W}\tilde{\eta}_j^2.$$

Следовательно, разбиение по дисперсии через одну точку (подобное) при отношении $n_A:n_B=1:1$ – это попытка получить подвыборки, связанные соотношением $tr(\chi_A)\approx tr(\chi_B)$. Если дисперсии на обеих подвыборках приблизительно одинаковы, то $tr(\chi_A)-tr(\chi_B)\approx 0$.

Максимально неподобное по дисперсии разбиение точек ряда, ранжированных по убыванию, получается следующим образом:

$$\max_{l=1,2}|tr(\mathbf{X}_{Al}^T\mathbf{X}_{Al})-tr(\mathbf{X}_{Bl}^T\mathbf{X}_{Bl})|=\max_{l=1,2}\left|\Delta_l^2\right|,$$

где l – вариант степени подобия.

Видно, что при отношении $n_A:n_B=2:1$ $(l=2)$ выражение $\left|\Delta_l^2\right|$ имеет бóльшее значение, чем при $n_A:n_B=1:1$ $(l=1)$, где n_A первых точек ряда имеют бóльшую дисперсию, т.е. $\Delta_2^2>\Delta_1^2$. Таким образом, при $l=2$ множества подвыборок более неподобные (несбалансированные по расстоянию между множествами), чем при $l=1$ и $\max_{l=1,2}\left|\Delta_l^2\right|=\Delta_2^2$.

При этом, однако, следует иметь в виду, что условие $tr(\mathbf{X}_A^T\mathbf{X}_A)=tr(\mathbf{X}_B^T\mathbf{X}_B)$ есть необходимым, но не достаточным для квадратичной зависимости информационных матриц и нахождения истинной модели с наименьшей дисперсией на всей выборке. Способ разбиения исходных данных по величине дисперсии следует считать простым способом, когда невозможно применять другие (если, например, вычислительный ресурс ограничен). Рассмотрим

наилучшее разбиение, если для выбора модели применяется критерий регулярности.

Постановка задачи отбора моделей с помощью критерия регулярности по результату оптимального разбиения выборки

В [4] доказано, что для получения оптимального разбиения выборки, при котором на подвыборках *A* и *B* обеспечивается неизменность структуры *s* и минимум дисперсии ошибки выхода модели $\hat{\mathbf{y}}_G = \mathbf{X}_{Gs}\hat{\theta}_{As}$, усредненной по всем реализациям шума, необходимо создать или соблюдать условие ρ^2-пропорциональности информационных матриц (1). Минимум матожидания дисперсии ошибки выхода – так называемого идеального критерия, имеет вид:

$$\overline{J}_G(s) = \min_s \mathrm{M}\left[\left\| \mathbf{y}_G^0 - \hat{\mathbf{y}}_{Gs} \right\|^2\right],$$
$$G = A, B,\ \mathrm{W} = \mathrm{A}\cup\mathrm{B},\ \mathrm{A}\cap\mathrm{B} = \varnothing, \qquad (2)$$

где M[·] обозначает символ математического ожидания; $\mathbf{y}_G^0$ – вектор выхода истинной модели, при этом G может быть любой из подвыборок *A* или *B*; $\hat{\mathbf{y}}_{Gs}$ – оценка выхода для зашумленной выборки G с использованием оценок, полученных методом наименьших квадратов (МНК) на части или на всей выборке. Оптимальная структура модели определяется как

$$s_J^* = \arg\min_{s=\overline{1,m}} \overline{J}_G(s). \qquad (3)$$

Критерий регулярности *AR*(*s*) отличается от идеального критерия (2) тем, что он вместо значений $\mathbf{y}_G^0$ использует реальные (зашумленные) значения выхода модели $\mathbf{y}_G = \mathbf{y}_G^0 + \xi_G$, и оценки параметров модели получаются по МНК на альтернативной части выборки, которая не участвует в оценивании выходного сигнала и критерия.

Формулировка видов критерия разбиения реальных данных

Пусть при моделировании объекта предполагается, что 1) «идеальная» модель существует, т.е. $\mathbf{y}^0 = \mathbf{X}^0\Theta^0$; 2) присутствует некоррелированный шум ξ на выходе $\mathbf{y} = \mathbf{y}^0 + \xi$, с нулевым математическим ожиданием, ограниченной дисперсией $\sigma^2 < \infty$ и диагональной ковариационной матрицей $\xi\xi^T = \sigma^2\mathbf{I}$, где $\mathbf{I}$ – единичная матрица. Предполагается также *некоррелированность* различных реализаций шума между собой и с полезным сигналом.

Пусть $E_J(\ell)$ – норма отклонения информационных матриц, соответствующих некоторому разбиению ℓ исходной матрицы на две подвыборки *A* и *B*. Критерий разбиения принадлежит множеству:

$$H_\upsilon = f[E_J(\ell)],\ \upsilon \in \Omega,\ J = \overline{1,5},\ \ell = \overline{1,L},$$

где L – множество всех возможных разбиений данных наблюдений на *A* и *B*; множество значений $J = \{1, 2, \ldots, 5\}$ для различного вида норм рассогласования информационных матриц (нормы представлены в [6]); Ω – множество видов критериев разбиения выборки; под $f \in \Xi$ понимается конкретный вид операции с нормой E_J из множества:

$$\Xi = \{\min(\|\cdot\|^2), \max(\|\cdot\|^2), \min(\|\cdot\|), \max(\|\cdot\|)\}, \qquad (4)$$

где $\|\cdot\|^2$ – обозначение квадратичной нормы, $\|\cdot\|$ – какая-либо другая норма (мера), например, обобщенная мера расстояний между подмножествами наблюдений, предложенная Г. Минковским. Здесь исследуются некоторые виды норм E_J: квадратичная и частный случай меры Минковского – *abs*(·), $J = \{1, 2\}$. Будем искать критерий H_υ (*CR*) для оптимального разбиения выборки

$$\ell^* = \arg\sup_{\ell=\overline{1,L}} f[E_J(\ell)],$$

соответствующий каждому критерию внешнего дополнения выбора лучшей модели при заданных $f \in \Xi$, $J = 1, 2$. В качестве внешнего критерия *CR* отбора моделей рассмотрим поочередно два критерия: регулярности *AR* и несмещенности решений $n_{\text{см}}^2$.

Анализ соответствия метода разбиения и критерия регулярности

Соотношение ρ^2-пропорциональности (1) должно выполняться для формирования ин-

формационных матриц обучающей A и проверочной B частей выборки, при условии обеспечения минимального значения идеального критерия для того, чтобы была получена единая истинная структура модели на всей выборке. Поскольку в практических задачах – это совместно трудновыполнимые условия, идеальный критерий заменяется критерием регулярности, а удовлетворение соотношения (1) – выполнением

$$\ell^* = \arg\min_{\ell=\overline{1,L}\ \rho_B^2\neq 0} \left\| \mathbf{X}_{A\ell}^T \mathbf{X}_{A\ell} - \rho_B^2 \mathbf{X}_{B\ell}^T \mathbf{X}_{B\ell} \right\| \qquad (5).$$

Если выполняется условие того, что оценки параметров модели определяются по данным подвыборки A, ($B\neq\varnothing$), $\mathbf{X}_A$ есть матрицей полностолбцового ранга $n_A \geq s$; на выходе присутствует аддитивный шум с выше заданными свойствами, то значения идеального критерия $J(S)$ и критерия регулярности $AR(s)$ будут различны. Рассмотрим четыре варианта исходных данных (вида матрицы $\mathbf{X}$ и величины дисперсии шума):

1. Все истинные аргументы присутствуют в матрице $\mathbf{X} = \mathbf{X}^0$, и дисперсия шума ξ в выходном векторе данных не превышает некоторый критический уровень $\sigma_{\text{кр}}^2$ ($0 \leq \sigma^2 < \sigma_{\text{кр}}^2$) [7].

2. Все «истинные» аргументы присутствуют $\mathbf{X} = \mathbf{X}^0$, но $\infty > \sigma^2 > \sigma_{\text{кр}}^2$.

3. Не все истинные аргументы присутствуют: в матрице $\mathbf{X}$: кроме части истинных $\mathbf{X}_{par}^0$ есть *ложные* $\mathbf{X}^{\xi}$, т.е. $\mathbf{X} = (\mathbf{X}_{par}^0 \vdots \mathbf{X}^{\xi})$, и шум имеет дисперсию $0 \leq \sigma^2 < \sigma_{\text{кр}}^2$.

4. В матрице $\mathbf{X} = (\mathbf{X}_{par}^0 \vdots \mathbf{X}^{\xi})$ присутствует шум с дисперсией $\infty > \sigma^2 > \sigma_{\text{кр}}^2$.

Тогда в первом варианте для сложности s определяемой структуры верно соотношение $s_J^* = s_{AR}^* = s^0$, где $s_{AR}^* = \arg\min_{s=\overline{1,m}} \overline{AR}(s)$. Во втором, третьем и четвертом вариантах $s_{AR}^* = s_J^* < s^0$.

Утверждение 1. Если все истинные аргументы присутствуют и дисперсия шума $0 \leq \sigma^2 < \sigma_{\text{кр}}^2$, то $s_J^* = s_{AR}^* = s^0$, но если $\infty > \sigma^2 > \sigma_{\text{кр}}^2$, тогда выполняется $s_{AR}^* = s_J^* < s^0$.

Сначала докажем отношение $s_J^* = s_{AR}^* = s^0$ в отсутствие шума (малом шуме), а потом $s_{AR}^* = s_J^* < s^0$ при наличии шума с дисперсией выше $\sigma_{\text{кр}}^2$. Доказательство следует из записи усредненных значений критериев

$$\overline{J}_{B|A} = \mathrm{M}\left[\left\| \mathbf{y}_B^0 - \mathbf{X}_B (\mathbf{X}_A^T \mathbf{X}_A)^{-1} \mathbf{y}_A^0 + \mathbf{X}_B (\mathbf{X}_A^T \mathbf{X}_A)^{-1} \xi_A \right\|^2 \right] =$$

$$= \left\| \mathbf{y}_B^0 - \mathbf{X}_B \bar{\Theta}_A \right\|^2 + \mathrm{M}\left[\left\| \mathbf{X}_B (\mathbf{X}_A^T \mathbf{X}_A)^{-1} \xi_A) \right\|^2 \right] =$$

$$= J_{B|A}^b + J_{B|A}^v ; \qquad (6)$$

$$\overline{AR}_{B|A} = \mathrm{M}\left[\left\| \mathbf{y}_B^0 - \mathbf{X}_B (\mathbf{X}_A^T \mathbf{X}_A)^{-1} \mathbf{y}_A^0 + \right.\right.$$

$$\left.\left. + (\xi_B - \mathbf{X}_B (\mathbf{X}_A^T \mathbf{X}_A)^{-1} \xi_A) \right\|^2 \right] = \left\| \mathbf{y}_B^0 - \mathbf{X}_B \bar{\Theta}_A \right\|^2 +$$

$$+ \mathrm{M}\left[\left\| (\xi_B - \mathbf{X}_B (\mathbf{X}_A^T \mathbf{X}_A)^{-1} \xi_A) \right\|^2 \right] = AR_{B|A}^b + AR_{B|A}^v , (7)$$

где $J_{B|A}^b$ и $AR_{B|A}^b$ – структурные, а $J_{B|A}^v$ и $AR_{B|A}^v$ – шумовые составляющие, причем $J_{B|A}^b = AR_{B|A}^b$.

При получении данных выражений было использовано условие некоррелированности полезного сигнала и шума и невырожденности информационных матриц. Оптимальная структура s_J^* находится минимизацией (3), а $s_{AR}^* = \arg\min_{s=\overline{1,m}} \overline{AR}(s)$.

Если шум равен нулю, структура модели находится минимизацией структурной составляющей, которая при увеличении s уменьшается до нуля. Структурная составляющая равна нулю, если все истинные аргументы входной матрицы включены в модель. Тогда $s^0 = \arg\min_s AR_{B|A}^b(s) = \arg\min_s J_{B|A}^b(s)$ или можно записать, что $\lim_{s\to s^0} AR_{B|A}^b(s) \to 0$ и $\lim_{s\to s^0} J_{B|A}^b(s) \to 0$. Тогда для истинной модели, имеющей структуру s^0, полученную и на A, и на B подвыборках, выполняются равенства $\mathbf{y}_A^0 = \mathbf{X}_A^0 \Theta_A^0$, $\mathbf{y}_B^0 = \mathbf{X}_B^0 \Theta_A^0$.

Рассмотрим шумовую составляющую подробнее. Если $AR_{B|A}^v \neq 0$, то из выражения (7) можно получить

$$AR^{v}_{A|B} = \sigma^2_B + \sigma^2_A tr(\mathbf{P}^T_{BA}\mathbf{P}_{BA}) \qquad (8)$$

и значение ее всегда больше, чем полученное из (6)

$$J^{v}_{A|B} = \sigma^2_A tr(\mathbf{P}^T_{BA}\mathbf{P}_{BA}). \qquad (9)$$

Здесь идемпотентная матрица

$$\mathbf{P}_{BA} = \mathbf{X}_B(\mathbf{X}^T_A\mathbf{X}_A)^{-1}\mathbf{X}^T_A \qquad (10)$$

имеет размерность $n_B \times n_A$. Покажем линейную зависимость (8) и (9) от сложности s на примере квадратной матрицы $\mathbf{X}_A$. Если записать шумовую составляющую для этого частного случая, то имеем

$$AR^{v}_{A|B} = \mathrm{M}[\,\xi^T_B\xi_B - 2(\mathbf{X}_B\mathbf{X}^{-1}_A\xi_A)^T\xi_B + \\ +\xi^T_A(\mathbf{X}_B\mathbf{X}^{-1}_A)^T(\mathbf{X}_B\mathbf{X}^{-1}_A)\xi_A\,]\,.$$

$$J^{v}_{A|B} = \mathrm{M}[\,2(\mathbf{X}_B\mathbf{X}^{-1}_A\xi_A)^T\xi_B + \xi^T_A(\mathbf{X}_B\mathbf{X}^{-1}_A)^T(\mathbf{X}_B\mathbf{X}^{-1}_A)\xi_A\,].$$

При получении данных выражений были использованы условия некоррелированности различных реализаций шума между собой $\mathrm{cov}(\xi^T_A\xi_B) = 0$, а также невырожденности матрицы $\mathbf{X}^{-1}_A$ ($\det\mathbf{X}_A \neq 0$). С учетом ρ-пропорциональности информационных матриц – равенства $\mathbf{X}^{-1}_B = \rho_B\mathbf{X}^{-1}_A$ при подстановке в выражение $\xi^T_A(\mathbf{X}_B\mathbf{X}^{-1}_A)^T(\mathbf{X}_B\mathbf{X}^{-1}_A)\xi_A$, получим $AR^{v}_{A|B} = \sigma^2_B + \sigma^2_A tr(\mathbf{I})/\rho^2_B$, $J^{v}_{A|B} = \sigma^2_A tr(\mathbf{I})/\rho^2_B$, где $\mathbf{I}$ – единичная матрица. Откуда видно, что след матрицы $tr(\mathbf{I})$, а также составляющие $AR^{v}_{A|B}$ и $J^{v}_{A|B}$ прямо пропорциональны сложности модели s, $AR^{v}_{A|B} = \sigma^2_B + s\sigma^2_A/\rho^2_B$, $J^{v}_{A|B} = s\sigma^2_A/\rho^2_B$ и, если дисперсии на обеих выборках одинаковы $\sigma^2 = \sigma^2_A = \sigma^2_B$, то $J^{v}_{A|B} = s\sigma^2/\rho^2_B = n\sigma^2/\rho^2_B$, $AR^{v}_{A|B} = \sigma^2(1 + \frac{1}{\rho^2_B}s) = \sigma^2(1 + \frac{1}{\rho^2_B}n)$. Видно, что минимум критериев $\bar{J}(s)$ и $\overline{AR}(s)$ при увеличении s достигается в одной точке $s^*_J = s^*_{AR}$, причем при любых σ^2 и s, так как $AR^{v}_{A|B}$ и $J^{v}_{A|B}$ имеют одинаковый коэффициент пропорциональности при s.

Аналогично для выражений (8) и (9), в которых используются прямоугольные матрицы $\mathbf{X}_A$ и $\mathbf{X}_B$, можно получить линейные зависимости от сложности s, вывод их не приводится здесь из-за громоздкости.

Таким образом, при нулевой или малой шумовой составляющей $J^{v}_{B|A}$ или $AR^{v}_{B|A}$, когда $0 \le \sigma^2 < \sigma^2_{\text{кр}}$, где $\sigma^2_{\text{кр}}$ соответствует шуму, при котором происходит упрощение истинной структуры модели s^0, утверждение $s^*_J = s^*_{AR} = s^0$ доказано.

Поскольку след матрицы $\mathbf{I}_{s=n}$ и $\left(P^T_{BA}P_{BA}\right)_{s<n}$ всегда положительный и с увеличением s растет, шумовые составляющие $AR^{v}_{A|B}$ и $J^{v}_{A|B}$ (в обоих случаях ρ- и ρ^2-пропорциональности данных) также растут, а структурные $AR^{b}_{B|A} = J^{b}_{B|A}$ – падают, то минимумы $\overline{AR}_{B|A}$ и $\bar{J}^{v}_{A|B}$ достигаются всегда раньше, чем минимум $AR^{b}_{B|A}$, если $\infty > \sigma^2 > \sigma^2_{\text{кр}}$. Так как $\sigma^2_{\text{кр}}$ соответствует шуму, при котором истинная структура модели теряет свою наименее коррелированную составляющую и становится более простой, то утверждение $s^*_J = s^*_{AR} < s^0$ доказано. По мере дальнейшего увеличения дисперсии шума происходит потеря последующих наименее коррелированных составляющих до тех пор, пока модель не дойдет до тривиальной, которой является константа.

Третий и четвертый варианты, когда истинные аргументы присутствуют не в полном составе и имеются «ложные», выполняется соотношение $s^*_{ARpar} = s^*_{Jpar} < s^0$. Доказательство очевидное. При любом уровне некоррелированного аддитивного шума минимум внешнего критерия достигается в точке соответствующей числу присутствующих истинных аргументов $s^*_{Jpar} = s^*_{ARpar}$, с увеличением шума отношение $s^*_{ARpar} = s^*_{Jpar} = s^*_{par}$ сохраняется. Если уровень шума нулевой или не превышающий первый критический уровень, то разница сложностей ис-

тинной структуры и полученной равна $s^0 - s^*_{par}$. С увеличением шума $\infty > \sigma^2 > \sigma^2_{\text{кр}}$ минимум сдвигается в сторону более простых моделей, разница увеличивается, получается модель со структурой $s^* < s^*_{par} < s^0$.

В связи с этим рассмотрим множество критериев неточного соблюдения равенства (1) и свойства норм отклонений от равенств. Были выделены две группы норм (4), которые соответствуют свойствам меры.

Если известна одна из подвыборок, например обучающая, можно спланировать эксперимент, создав проверочную выборку согласно соотношению (1). В случае пассивного эксперимента к выполнению условия (1) можно приблизиться, минимизируя норму рассогласования $\left\| \mathbf{X}^T_{A\ell}\mathbf{X}_{A\ell} - \rho^2_B \mathbf{X}^T_{B\ell}\mathbf{X}_{B\ell} \right\|$ путем выбора разбиения ℓ^*.

Поскольку среднее значение критерия регулярности $\overline{AR}(\cdot)$ является оценкой теоретической дисперсии ошибки модели $\bar{J}_B(s)$, то разбиение, удовлетворяющее (1), в пределе по множеству усреднений результатов численных экспериментов справедливо и для него, с учетом неизменности структур на множестве реализаций шума в выходных данных. Тогда существует структура

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

где s^* – оптимальная структура при разбиении, удовлетворяющем (5).

Таким образом, при выборе моделей по минимуму критерия регулярности следует минимизировать критерий нормы квазиоптимального разбиения либо, в силу большого перебора, применять вычислительно менее затратное *подобное по дисперсии* разбиение, являющееся частным случаем квазиоптимального.

Выбор модели по критерию несмещенности решений при квазиоптимальном разбиении выборки

Пусть выполняется условие ρ^2-пропорциональности, и структуры моделей на различных частях выборки идентичны, т.е. $s_A = s_B$. После выбора оптимальной структуры оценки векторов коэффициентов на различных подвыборках неодинаковы, т.е. $\hat{\theta}_A \neq \hat{\theta}_B$. Критерий несмещенности (смещения решения) имеет вид:

$$n^2_{см}(s) = \left\| \hat{\mathbf{y}}_W(\hat{\theta}_{As}) - \hat{\mathbf{y}}_W(\hat{\theta}_{Bs}) \right\|^2 = \left\| \mathbf{X}_{Ws}(\hat{\Theta}_{As} - \hat{\Theta}_{Bs}) \right\|^2 =$$
$$= (\hat{\Theta}_{As} - \hat{\Theta}_{Bs})^T \mathbf{X}^T_{Ws}\mathbf{X}_{Ws}(\hat{\Theta}_{As} - \hat{\Theta}_{Bs}).$$

Этот критерий можно представить в виде суммы структурной и шумовой составляющих

$$n^2_{\text{см}}(s) = n^{2\ b}_{\text{см}}(s) + n^{2\ v}_{\text{см}}(s).$$

Структурную составляющую этого критерия с учетом матрицы канонической формы $\mathbf{D}(s)$ [8] можно записать как:

$$n^{2\ b}_{см}(s) = \left(\mathbf{y}^0\right)^T \mathbf{D}(s)\mathbf{y}^0 =$$
$$= \left((\mathbf{y}^0_A)^T, (\mathbf{y}^0_B)^T\right) \begin{pmatrix} \mathbf{P}^T_{WAs}\mathbf{P}_{WAs} & -\mathbf{P}^T_{WAs}\mathbf{P}_{WBs} \\ -\mathbf{P}^T_{WBs}\mathbf{P}_{WAs} & \mathbf{P}^T_{WBs}\mathbf{P}_{WBs} \end{pmatrix} \begin{pmatrix} \mathbf{y}^0_A \\ \mathbf{y}^0_B \end{pmatrix},$$

где идемпотентная матрица $\mathbf{P}$ определяется аналогично формуле (10), в которой индекс s опущен. Если использовалось условие ρ^2-пропорциональности выборок и того, что матрицы $\mathbf{X}_A$ и $\mathbf{X}_B$ есть матрицами полностолбцового ранга, то полученный результат можно преобразовать к виду:

$$\begin{aligned} n^{2\ b}_{см}(s) &= (1+\frac{1}{\rho^2})\left\{[(\mathbf{y}^0_A)^T\mathbf{X}_{As} - \rho^2(\mathbf{y}^0_B)^T\mathbf{X}_{Bs}]\times\right. \\ &\left.\times(\mathbf{X}^T_{As}\mathbf{X}_{As})^{-1}[\mathbf{X}^T_{As}\mathbf{y}^0_A - \rho^2\mathbf{X}^T_{Bs}\mathbf{y}^0_B]\right\} = \\ &= \frac{1}{\rho^2}(1+\frac{1}{\rho^2})\left\{[\mathbf{X}^T_{As}\mathbf{y}^0_A - \rho^2\mathbf{X}^T_{Bs}\mathbf{y}^0_B]^T\times\right. \\ &\left.\times(\mathbf{X}^T_{Bs}\mathbf{X}_{Bs})^{-1}[\mathbf{X}^T_{As}\mathbf{y}^0_A - \rho^2\mathbf{X}^T_{Bs}\mathbf{y}^0_B]\right\}. \end{aligned} \qquad (11)$$

Если в выборке W присутствует неполный набор истинных аргументов и $s^0 > s^0_{par}$ и выполняется условие ρ^2-пропорциональности при отсутствии шума (малом шуме), т.е. $0 \le \sigma^2 < \sigma^2_{кр}$, то минимум внешнего критерия достигается в точке, соответствующей числу присутствующих истинных аргументов $s^*_{Jpar} = s^*_{ARpar} = s^0_{par}$. Ввиду выполнения $\mathbf{X}^0_{Apar}\mathbf{y}^0_A = \rho^2\mathbf{X}^0_{Bpar}\mathbf{y}^0_B$ и присутствия ложных входных переменных оценки коэффициентов векторов

неодинаковы ($\widehat{\Theta}_A \neq \widehat{\Theta}_B$), так как $n_A \neq n_B$. При этом также предполагается, что ложные входные переменные не коррелированы с выходом и между собой, т.е. являются шумом.

Структурная составляющая (11) тождественно равна нулю при наличии:

- в матрице полностолбцового ранга полного набора истинных аргументов, ($\mathbf{X} = \mathbf{X}^0$) (т.к. $\mathbf{y}_B^0 = \mathbf{X}_B^0\Theta^0$, $\mathbf{y}_A^0 = \mathbf{X}_A^0\Theta^0$) и ρ^2-пропорциональности данных $\forall\, s^0 \le s < n_A \neq n_B$;
- всех истинных аргументов $\mathbf{X} = \mathbf{X}^0$ и $s = n_A = n_B$ ρ-пропорциональности данных.

Тогда $s^* = s^0 = s_A^o = s_B^o$, $\widehat{\Theta}_{As^*} = \widehat{\Theta}_{Bs^*} = \Theta^0$. Когда в модели присутствует неполный набор истинных аргументов (т.е. $\widehat{\Theta}_A \neq \widehat{\Theta}_B \neq \Theta^0$), то с ростом числа *s* структурная составляющая, если не создать условия целесообразного процесса построения моделей, изменяется произвольно, и имеет несколько локальных минимумов.

Рассмотрим шумовую составляющую $n_{\text{см}}^{2\,v}(s)$ после усреднения на множестве выборок при условии ρ^2-пропорциональности. Рассмотрим среднее значение второй составляющей критерия несмещенности решений, если дисперсия на выборках *A* и *B* одинакова:

$$n_{\text{см}}^{2\,v}(s) = \mathrm{M}(\xi^T\mathbf{D}(s)\xi) = \sigma^2[tr(\mathbf{P}_{WAs}^T\mathbf{P}_{WAs} + \mathbf{P}_{WBs}^T\mathbf{P}_{WBs})] =$$
$$= \sigma^2[tr(\mathbf{P}_{WAs}^T\mathbf{P}_{WAs}) + tr(\mathbf{P}_{WAs}^T\mathbf{P}_{WAs})] =$$
$$= \sigma^2[tr\{(\mathbf{X}_{As}^T\mathbf{X}_{As})^{-1}\mathbf{X}_{Ws}^T\mathbf{X}_{Ws}\} + tr\{(\mathbf{X}_{Bs}^T\mathbf{X}_{Bs})^{-1}\mathbf{X}_{Ws}^T\mathbf{X}_{Ws}\}] =$$
$$= \sigma^2 tr\{[(\mathbf{X}_{As}^T\mathbf{X}_{As})^{-1} + (\mathbf{X}_{Bs}^T\mathbf{X}_{Bs})^{-1}]\mathbf{X}_{Ws}^T\mathbf{X}_{Ws}\}.$$

При условии ρ^2-пропорциональности и того, что матрицы $\mathbf{X}_{As}$ и $\mathbf{X}_{Bs}$ являются матрицами полного ранга, результат можно преобразовать к виду

$$n_{\text{см}}^{2\,v}(s) = \sigma^2[tr\{(1+\rho_B^2)(\mathbf{X}_{As}^T\mathbf{X}_{As})^{-1}\mathbf{X}_{Ws}^T\mathbf{X}_{Ws}\} =$$
$$= \sigma^2(1+\rho_B^2)tr\{(\mathbf{X}_{As}^T\mathbf{X}_{As})^{-1}(\mathbf{X}_{As}^T\mathbf{X}_{As} + \mathbf{X}_{Bs}^T\mathbf{X}_{Bs})\}$$
$$= \sigma^2 s(1+\rho_B^2)^2/\rho_B^2.$$

Шумовая составляющая с ростом сложности структуры *s* линейно растет при одинаковой дисперсии шума σ^2. Если $\sigma_A^2 \neq \sigma_B^2 \neq \sigma^2$, то

$$n_{\text{см}}^{2\,v}(s) = s(1+\rho_B^2)(\sigma_B^2 + \sigma_A^2/\rho_B^2).$$

Если выполняются условия ρ-пропорциональности и неравенства дисперсий $\sigma_A^2 \neq \sigma_B^2$, то шумовая составляющая

$$\tilde{n}_{\text{см}}^{2\,v}(s) = \sigma_A^2 tr\{(\mathbf{X}_A)^{-1}\mathbf{X}_W\} + \sigma_B^2 tr\{(\mathbf{X}_B)^{-1}\mathbf{X}_W\} =$$
$$= \sigma_A^2 tr\{(\mathbf{X}_A)^{-1}\mathbf{X}_A + (\mathbf{X}_A)^{-1}\mathbf{X}_B\} +$$
$$+ \sigma_B^2 tr\{(\mathbf{X}_B)^{-1}\mathbf{X}_A + (\mathbf{X}_B)^{-1}\mathbf{X}_B\} =$$
$$= \sigma_A^2 s(1+\frac{1}{\rho_B}) + \sigma_B^2 s(1+\rho_B) =$$
$$= s\frac{(1+\rho_B)}{\rho_B}(\sigma_A^2 + \rho_B\sigma_B^2)$$

также линейно зависит при возрастании числа *s*.

Если $\sigma^2 = \sigma_A^2 = \sigma_B^2$, то $\tilde{n}_{\text{см}}^{2\,v}(s) = s\sigma^2\frac{(1+\rho_B)^2}{\rho_B}$.

Если $\rho_B < 0$, то шумовая составляющая положительна. Тогда критерий несмещенности решений будет иметь минимум в положительной области, если $\rho_B < 0$ – то в отрицательной области значений, когда $\mathbf{X} = \mathbf{X}^0$ или $\mathbf{X} = \mathbf{X}_{par}^0$ шум некоррелированный, а также нулевой или небольшой, т.е. $0 \le \sigma^2 < \sigma_{\text{кр}}^2$.

Если данные подвыборок ρ^2-пропорциональны, ρ-*пропорциональны* или получены квазиоптимальным разбиением, то критерий несмещенности решений в общем случае не будет адекватен шуму. Так как если присутствуют все истинные аргументы и некоррелированный аддитивный шум, который растет, то при минимизации $n_{\text{см}}^2(s)$ выбирается тривиальная модель минимальной (нулевой) структуры.

Для того чтобы он был адекватным его следует либо вообще не применять в вышеуказанных случаях, либо применять при максимизации квадратичной нормы рассогласования информационных матриц.

В рассмотренных случаях структурная составляющая внешнего критерия всегда имеет минимум и этот минимум соответствует на оси сложности структур случаю равенства структур и параметров моделей, полученных на вы-

борках A и B для истинной структуры, т.е. если $\mathbf{X} = \mathbf{X}^0$. Поэтому после максимизации квадратичной нормы рассогласования информационных матриц необходимо минимизировать значение критерия несмещенности.

Если в данных присутствуют все истинные аргументы и малый шум с дисперсией $0 \le \sigma^2 < \sigma^2_{\text{кр}}$, а для разбиения выборки максимизируется квадратичная норма рассогласования информационных матриц, то глобальный минимум критерия несмещенности решений соответствует частичной истинной модели, включающей все истинные аргументы из матрицы $\mathbf{X} = \mathbf{X}^0_{par}$.

С увеличением дисперсии шума $\infty > \sigma^2 > \sigma^2_{\text{кр}}$ минимум будет сдвигаться в сторону все более простых моделей, но по абсолютной величине он будет изменяться произвольно.

Для того, чтобы структурная составляющая имела вид ниспадающей кривой, нужно, чтобы разбиение выборки обеспечивало этот процесс изменения. В регрессионном методе включения на каждом шаге в модель вводятся аргументы, вносящие наибольший вклад в уменьшение ошибки модели. При условии нормирования данных, как правило, это – имеющие наибольшие по модулю коэффициенты. Чтобы разницы оценок коэффициентов на различных выборках – составная часть критерия несмещенности решений, при одном и том же числе s была наибольшая, нужно, чтобы выборки были максимально неподобные. Поэтому строить модели с применением критерия несмещенности решений предлагается по данным максимально различающихся подвыборок наблюдений объекта. Например, в качестве критерия разбиения использовать не минимум нормы $\left\| \mathbf{X}^T_{A\ell}\mathbf{X}_{A\ell} - \rho^2_B \mathbf{X}^T_{B\ell}\mathbf{X}_{B\ell} \right\|$, а ее максимум. Тогда наилучшее разбиение ℓ^* будет иметь вид

$$\ell^* = \arg \max_{\rho^2_{B\ell} \ne 0,\ \ell=\overline{1,L}} \left\| \mathbf{X}^T_{A\ell}\mathbf{X}_{A\ell} - \rho^2_{B\ell} \mathbf{X}^T_{B\ell}\mathbf{X}_{B\ell} \right\|.$$

Либо для получения разбиения ℓ^* использовать неподобные по дисперсии подвыборки A и B

$$l^* = \arg \max_{l=\overline{1,L_d}} tr(\tilde{\mathbf{X}}^T_{Al}\tilde{\mathbf{X}}_{Al} - \tilde{\mathbf{X}}^T_{Bl}\tilde{\mathbf{X}}_{Bl}) =$$

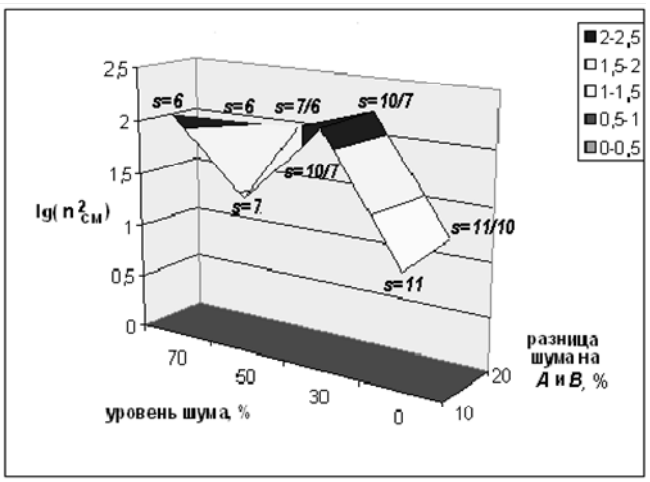
$$= \arg \max_{l=\overline{1,L_d}} \left[tr(\tilde{\mathbf{X}}^T_{Al}\tilde{\mathbf{X}}_{Al}) - tr(\tilde{\mathbf{X}}^T_{Bl}\tilde{\mathbf{X}}_{Bl}) \right],$$

где для вычисления дисперсий используются диагональные элементы информационных матриц, элементы которых центрированы как: $\tilde{x}_{ij} = x_{ij} - \bar{x}_j$, $j = \overline{1,m}$, $i = \overline{1,n_W}$, а $\bar{x}_j$ – средние значения аргументов, $j = \overline{1,m}$, вычислены по данным матрицы $\mathbf{X}_W$. Индекс l используется для обозначения набора строк при переборе содержания матриц $\mathbf{X}_{Al}$ и $\mathbf{X}_{Bl}$, $\mathbf{X}_W = \mathbf{X}_{Al} \cup \mathbf{X}_{Bl}$.

Результаты численного эксперимента

Пусть имеется матрица исходных данных $\left(\mathbf{X}^0 : \mathbf{y}\right)$ размерности $n \times (m+2)$, содержащая все истинные аргументы, где на выходе имеется помеха $y = y^0 + \xi$. Пусть шум генерируется при помощи генератора псевдослучайных чисел с равномерным законом $\zeta \in [0;1]$, принимает в процентном отношении к величине размаха изменения выходной величины $(y^0_{\max} - y^0_{\min})$ значения $\xi = \alpha(\zeta - 1)(y^0_{\max} - y^0_{\min})/200$, где α – уровень шума в процентах. Проведем эксперимент, заключающийся в проверке того, как изменятся критерий несмещенности решений, если нужно восстановить истинную структуру модели $y^0 = \theta_0 + \sum_{i=1}^{m} \theta_i x^0_i$, линейную по десяти входным переменным и одиннадцати коэффициентам, сгенерированным по равномерному случайному закону распределения на интервале $-\beta \le \theta_i \le \beta$, $i = \overline{0,m}$, содержащую $m+1 = 11$ коэффициентов. Сложность структуры модели $s^0 = 11$ с ненулевыми коэффициентами и модель содержит свободный член. Переменные также принимают случайные значения на интервале [0;1]. Проследим, как изменяется сложность моделей, выбираемых по минимуму критерия несмещенности решений. Исследуем результат моделирования при изменении минимальных уровней шума и различных отклонениях дисперсий в подвыборках A и B.

Проанализируем результаты для двух значений 10 и 20 процентов отклонений от минимальных значений шума одной из подвыборок, принимающих значения α = [0, 30, 50, 70] процентов. На рисунке приведены результаты этого численного эксперимента, где видно, как упрощаются модели, т.е. какие-то коэффициенты обнуляются, а с ростом шума, с меньшей разницей шума это происходит менее предсказуемо, чем с бóльшей. Чем больше разница шума в выборках, тем кривая изменения $\lg(n^2_{\text{см}})$ более гладкая, а критерий – более адекватен шуму. В записи «$s = 10/7$» первая цифра означает сложность на подвыборке с меньшей дисперсией, вторая – на выборке с бóльшей.



Изменение минимума критерия несмещенности решений, соответствующего оптимальной модели, при увеличении уровня шума в выборке W для двух значений разницы уровней в подвыборках A и B (неподобных по дисперсии подвыборок)

Заключение. Неподобное по дисперсии разбиение целесообразно при использовании критерия несмещенности, причем лучший результат, следует ожидать при отношении $n_A : n_B$ = = 2:1, а не при отношении $n_A : n_B$ =1:1.

При выборе моделей по минимуму критерия несмещенности решений следует:

– избегать квадратичной и линейной пропорциональности данных, квазиоптимального разбиения, так же, как и производить «подобное по дисперсии» разбиение;

– использовать максимизацию нормы разности информационных матриц или «неподобные по дисперсии» разбиения выборки.

1. *Ивахненко А.Г.* Системы эвристической самоорганизации в технической кибернетике. – Киев: Техніка, 1971. – 372 с.
2. *Висоцький В.М.* Про найкращий поділ вихідних даних в алгоритмах МГУА // Автоматика. – 1976. – № 3. – С. 71–74.
3. *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.
4. *Степашко В.С.* Структурная идентификация прогнозирующих моделей в условиях планируемого эксперимента // Там же. – 1992. – № 1. – С. 26–35.
5. *Степашко В.С.*, *Кондрашова Н.В.* Исследование свойств критериев разбиения выборки в алгоритмах МГУА // АСУ і прогресивні інформаційні технології. – 2005. – **3**. – С. 116–123.
6. *Степашко В.С.*, *Кондрашова Н.В.* Анализ проблемы разбиения выборки для алгоритмов МГУА // Кибернетика и вычислительная техника. – 2002. – **136**. – С. 3–15.
7. *Степашко В.С.* Метод критических дисперсий как аналитический аппарат теории индуктивного моделирования // Проблемы управления и информатики. –2008. – № 2. – С. 8–26.
8. *Юрачковский Ю.П.*, *Грошков А.Н.* Применение канонических форм внешних критериев для исследования их свойств // Автоматика. – 1979. – № 3. – С. 85–89.

Поступила 10.11.2014

Тел. для справок: +38 044 526-3028 (Киев)
E-mail: *nkondrashova@ukr.net*

© Н.В. Кондрашова, 2015

UDC 681.513.8

Kondrashova N.V.

The Influence of Data Division on the Adequacy of the External Addition Criterion in Group Method of Data Handling Algorithms

The article is devoted to solving the so-called "problem of the partitioning" in the group method of handling arguments (GMDH).

The article is based on the results known in literature as the criterion of regularity. It investigates the problem of partitioning for GMDH criteria belonging to the class criterion of minimum bias (or unbiased criteria), namely, to criterion the un-

biasedness solutions. The geometric interpretation is given of the variance partitioning of samples traditionally used in different proportions of their volumes. Formulation of diversity of species partitioning criteria is represented. The statement of the problem of models selection using the criterion of regularity on the result of optimal partitioning samples is considered and analyzed the compliance of the method partitioning and the regularity criterion.

We investigate the validity of the joint application of the method data partitioning and criterion external additions in finding the most accurate models. Analysis of ρ-optimal partitionings and of their practicability analogue – quasi-optimal partitioning of samples is given. Conformity with the criteria of the external addition mentioned partitionings when selecting the best model structure discussed. Formulated what kind of selection criterion of models should be used with the particular method of partitioning the original sample and what view of at the same time should have criterion to partitioning.

Theoretical analysis substantiates the common using of maximizing the difference of dispersions in the subsamples partition and minimization of external criterion unbiasedness solutions. If the condition of proportionality of data was satisfied, the theoretical substantiation and confirmation in numerical experiments were obtained. The criterion of unbiasedness parameters, is not "adequate" criterion while minimizing the criterion of the sample division, but only when it is maximized. Numerical experiment at various levels of noise in the data indicates that such method partitioning contributes to finding the "true" arguments in the models in GMDH algorithms with minimization criterion unbiasedness solutions.

●