МЕТОД ГРУППОВОГО УЧЕТА АРГУМЕНТОВ

Метод группового учета аргументов, МГУА (Group Method of Data Handling, GMDH) — метод порождения и выбора регрессионных моделей оптимальной сложности. Под сложностью модели в МГУА понимается число параметров. Для порождения используется базовая модель, подмножество элементов которой должно входить в искомую модель. Для выбора моделей используются внешние критерии, специальные функционалы качества моделей, вычисленные на тестовой выборке.

МГУА рекомендуется к использованию в том случае, когда выборка содержит несколько элементов. Тогда при построении регрессионных моделей использовать статистические гипотезы о плотности распределения, например, гипотезу о Гауссовском распределении, невозможно. Поэтому используется индуктивный подход, согласно которому последовательно порождаются модели возрастающей сложности до тех пор, пока не будет найден минимум некоторого критерия качества модели. Этот критерий качества называется *внешний критерий*, так как при настройке моделей и при оценке качества моделей используются разные данные. Достижение глобального минимума внешнего критерия при порождении моделей означает, что модель, доставляющая такой минимум, является искомой.

Один из авторов этого метода А.Г. Ивахненко пишет: «Осуществляется целенаправленный перебор многих моделей-претендентов различной сложности по ряду критериев. В результате находится модель оптимальной структуры в виде одного уравнения или системы уравнений. Минимум критерия селекции определяет модель оптимальной структуры».

Индуктивный алгоритм отыскания модели оптимальной структуры состоит из следующих основных шагов.

1. Пусть задана выборка

$$D=\left\{(x_n,y_n)\right\}_{n=1}^N,\ x\in\Re^m. \quad (1.1)$$

Выборка разбивается на обучающую и тестовую. Обозначим ℓ,C — множества индексов из $\{1,...,N\}=W$. Эти множества удовлетворяют условиям разбиения $\ell\cup C=W,\ell\cap C=\emptyset$. Матрица X_ℓ состоит из тех векторов-строк x_n, для которых индекс $n\in\ell$. Вектор y_ℓ состоит из тех элементов y_n, для которых индекс $n\in\ell$. Разбиение выборки представляется в виде

2. Назначается базовая модель. Эта модель описывает отношение между зависимой переменной y и свободными переменными x. Например, используется функциональный ряд Вольтерра, называемый также полиномом Колмогорова-Габора:

$$y=w_0+\sum_{i=1}^m w_i x_i+\sum_{i=1}^m\sum_{j=1}^m w_{ij}x_i x_j+\sum_{i=1}^m\sum_{j=1}^m\sum_{k=1}^m w_{ijk}x_i x_j x_k+\ldots. \quad (1.2)$$

В этой модели $x=\left\{x_i|i=1,...,m\right\}$ — множество свободных переменных и w — вектор параметров — весовых коэффициентов

$w=\langle w_i,w_{ij},w_{ijk},...|i,j,k,...=1,...,m\rangle$.

В некоторых случаях имеет смысл увеличить число элементов вектора свободной переменной x за счет добавления нелинейных преобразований отдельных переменных. Например, задано конечное множество нелинейных функций $G=\left\{g|\Re\longrightarrow\Re\right\}$. Дополнительная свободная переменная получается путем применения некоторого преобразования из G к одной или к нескольким переменным из множества $\{x\}$. Базовая модель линейна относительно параметров w и нелинейна относительно свободных переменных x.

3. Исходя из поставленных задач, выбирается целевая функция — внешний критерий, описывающий качество модели. Ниже описаны несколько часто используемых внешних критериев.
4. Индуктивно порождаются модели-претенденты. При этом вводится ограничение на длину полинома базовой модели. Например, степень полинома базовой модели на не должно превышать заданное число R. Тогда базовая модель представима в виде линейной комбинации заданного числа F_0 произведений свободных переменных:

$$y = f(x_1, x_2, \ldots, x_1^2, x_1 x_2, x_2^2, \ldots, x_m^R), \quad (1.3)$$

здесь f — линейная комбинация. Аргументы этой функции переобозначаются следующим образом:

$$x_1 \to a_1, x_2 \to a_2, \ldots, x_1^2 \to a_\alpha, x_1 x_2 \to a_\beta, x_2^2 \to a_\gamma, \ldots, x_m^q \to a_{F_0}, \quad (1.4)$$

то есть,

$$y = f(a_1, a_2, \ldots, a_{F_0}). \quad (1.5)$$

Для линейно входящих коэффициентов задается одноиндексная нумерация $w = \langle w_1, \ldots, w_{F_0} \rangle$. Тогда модель может быть представлена в виде линейной комбинации

$$y = w_0 + \sum_{i=1}^{F_0} w_i a_i = w_0 + \mathbf{w} \cdot \mathbf{a}$$

— скалярное произведение. (1.6)

Каждая порождаемая модель задается линейной комбинацией элементов $\{(w_i, a_i)\}$, в которой множество индексов $\{i\} = s$ является подмножеством $\{1, \ldots, F_0\}$.

5. Настраиваются параметры моделей. Для настройки используется *внутренний критерий* — критерий, вычисляемый с использованием обучающей выборки. Каждому элементу вектора x_n — элемента выборки D ставится в соответствие вектор a_n, алгоритм построения соответствия указан выше. Строится матрица A_W — набор векторов-столбцов a_i. Матрица A_W разбивается на подматрицы A_ℓ и A_C. Наименьшую невязку $|y - \hat{y}|$, где

доставляет значение вектора параметров $\widehat{w}$, который вычисляется методом наименьших квадратов:

$$\widehat{w}_G=(A_G^T A_G)^{-1} A_G^T y_G, \text{ (1.7), где } G\in\{\ell,C,W\}.$$

При этом в качестве внутреннего критерия выступает среднеквадратичная ошибка

$$\epsilon_G^2=|y_G-A_G\widehat{w}_G|^2. \text{ (1.8)}$$

В соответствии с критерием $\epsilon_G^2 \longrightarrow \min$ происходит настройка параметров w и вычисление ошибки на тестовой подвыборке, обозначенной G, здесь $G=\ell$. При усложнении модели внутренний критерий не дает минимума для моделей оптимальной сложности, поэтому для выбора модели он не пригоден.

6. Для выбора моделей вычисляется качество порожденных моделей. При этом используется контрольная выборка и назначенный внешний критерий. Ошибка на подвыборке H обозначается

$$\Delta^2(H)=\Delta^2(H\backslash G)=|y_H-A_H\widehat{w}_G|^2, \text{ (1.9),}$$

где $H\in\{\ell,C\}$, $H\cap G=\emptyset$. Это означает, что ошибка вычисляется на подвыборке H при параметрах модели, полученных на подвыборке G.

7. Модель, доставляющая минимум внешнему критерию, считается оптимальной.

Если значение внешнего критерия не достигает своего минимума при увеличении сложности модели или значение функции качества неудовлетворительно, то выбирается лучшая модель из моделей заданной сложности. Под сложностью модели подразумевается число настраиваемых параметров модели. Существуют следующие причины, по которым глобальный минимум может не существовать:

данные слишком зашумлены,

среди данных нет необходимых для отыскания модели переменных,

неверно задан критерий выбора,

при анализе временных рядов существует значительная временная задержка отыскиваемой причинно-следственной связи.

1.1. Внешние критерии МГУА

Авторами метода рассмотрено весьма большое число различных критериев выбора моделей. Значительная часть этих критериев опубликована на сайте http://www.gmdh.net.

Критерий выбора модели может быть назван внешним, если он получен с помощью дополнительной информации, не содержащейся в данных которые использовались при вычислении параметров моделей. Например, такая информация содержится в дополнительной, тестовой выборке.

Алгоритм МГУА использует и внутренний критерий и внешний. Внутренний критерий используется для настройки параметров модели, внешний критерий используется для выбора модели оптимальной структуры. Возможен выбор моделей по нескольким внешним критериям.

1.1.1. Критерий регулярности

Критерий регулярности $\Delta^2(C)$ включает среднеквадратичную ошибку на обучающей подвыборке C, полученную при параметрах модели, настроенных на тестовой подвыборке ℓ.

$$\Delta^2(C) = |y_C - A_C \hat{w}_\ell|^2 = (y_C - A_C \hat{w}_\ell)^T (y_C - A_C \hat{w}_\ell), \quad (1.1.1.1)$$

где

$$\hat{w}_\ell = (A_\ell^T A_\ell)^{-1}(A_\ell^T y_\ell) \text{ и } \hat{y}_C(\ell) = A_C \hat{w}_\ell.$$

Другие модификации критерия регулярности:

$$\Delta^2(C) = \frac{|y_C - A_C \hat{w}_\ell|^2}{|y_C|^2} \quad (1.1.1.2)$$

и

$$\Delta^2(C) = \frac{|y_C - A_C \hat{w}_\ell|^2}{|y_C - \bar{y}_C|^2}, \quad (1.1.1.3)$$

где $\bar{y}$ — среднее значение вектора y.

Критерий $\Delta^2(C)$ также обозначается $\Delta^2(C \backslash \ell)$, то есть ошибка на подвыборке C, при параметрах, полученных на подвыборке ℓ.

1.1.2. Критерий минимального смещения

Иначе критерий непротиворечивости модели: модель которая имеет на обучающей выборке одну невязку, а на контрольной — другую, называется противоречивой. Этот критерий включает разность между зависимыми переменными модели, вычисленными на двух различных выборках ℓ и C. Критерий не включает ошибку модели в явной форме. Он требует, чтобы оценки коэффициентов в оптимальной модели, вычисленные на множествах ℓ и C, различались минимально.

Критерий непротиворечивости как критерий минимума смещения имеет вид:

$$\eta_{\text{bs}}^2 = |A_W \widehat{w}_\ell - A_W \widehat{w}_C|^2 = (\widehat{w}_\ell - \widehat{w}_C)^T A_W^T A_W (\widehat{w}_\ell - \widehat{w}_C). \quad (1.1.2.1)$$

Другие модификации этого критерия:

$$\eta_{\text{bs}}^2 = \frac{|A_W \widehat{w}_\ell - A_W \widehat{w}_C|^2}{|y_C - \bar{y}_C|^2} \quad (1.1.2.2)$$

и

$$\eta_{\text{a}}^2 = |\widehat{w}_\ell - \widehat{w}_C|^2, \quad (1.1.2.3)$$

где $\widehat{w}_\ell$ и $\widehat{w}_C$ — векторы коэффициентов, полученные с использованием подвыборок ℓ и C. При использовании последнего варианта следует помнить, что число элементов вектора параметров w в различных моделях может быть различно.

1.1.3. Критерий absolute noise-immune

Утверждается, что с помощью этого критерия, из сильно зашумленных данных возможно найти скрытые физические закономерности.

$$V^2 = (A_W \widehat{w}_\ell - A_W \widehat{w}_W)^T (A_W \widehat{w}_W - A_W \widehat{w}_C) =$$
$$= (\widehat{w}_\ell - \widehat{w}_W)^T A_W^T A_W (\widehat{w}_W - \widehat{w}_C). \quad (1.1.3.1)$$

где $\widehat{w}_W$ — вектор коэффициентов, полученный на всей выборке W.

1.1.4. Критерий предсказательной способности

Является модификацией критерия регулярности. Этот критерий включает среднеквадратичную ошибку для отдельной экзаменационной выборки B, которая не была использована ни при нахождении коэффициентов, ни при выборе моделей. В этом случае выборка делится не на две, а на три части:

Критерий предсказательной способности имеет вид:

$$\Delta^2(W\setminus B)=\frac{|y_W-A_W\hat{\omega}_B|^2}{|y_W-\bar{y}_W|^2}. \quad (1.1.4.1)$$

1.1.5. Комбинированный критерий

Этот критерий позволяет использовать при выборе моделей линейную комбинацию нескольких критериев. Комбинированный критерий $k^2=\sum_{i=1}^{K}\alpha_i k_i^2,$ при условии нормировки $\sum_{i=1}^{K}\alpha_i=1.$

Здесь k_i — принятые на рассмотрение критерии, а α_i — веса этих критериев, назначенные в начале вычислительного эксперимента. Используются также нормализованные значения критериев. При этом предыдущая формула имеет вид

$$k^2=\sum_{i=1}^{K}\alpha_i\frac{k_i^2}{k_{i\max}^2}. \quad (1.1.5.1)$$

Максимальное значение критерия $k_{i\max}^2$ берется по вычисленным значениям критериев для всех порожденных моделей. В данном случае оптимальная модель может быть найдена только после завершения настройки параметров всех моделей.

Пример распространенного комбинированного критерия — смещение плюс ошибка аппроксимации.

$$c_1^2=\bar{\eta}_{bs}^2+\bar{\epsilon}^2(W)=\frac{\eta bs^2}{\eta bsmax^2}+\frac{\epsilon^2}{\epsilon_{\max}^2}, \quad (1.1.5.2)$$

где $\bar{\epsilon}^2(W)$ — нормализованная среднеквадратичная ошибка аппроксимации на всей выборке $W=\ell\cup C$ с использованием коэффициентов, полученных также на W.

Второй пример комбинированного критерия — смещение плюс регулярность.

$$c_2^2 = \bar{\eta} b s^2 + \overline{\Delta}^2(C). \quad (1.1.5.3)$$

Третий пример — смещение плюс ошибка на тестовой выборке.

$$c_3^2 = \bar{\eta} b s^2 + \overline{\Delta}^2(B \backslash W). \quad (1.1.5.4)$$

Такой критерий обеспечивает выбор наиболее несмещенных, устойчивых и точных моделей. Здесь $\Delta(C \backslash W)$ — среднеквадратичная ошибка, вычисленная на выборке C, с весами, настроенными на всей выборке W.

Обычно при вычислении критерия c_3 выборку делят на три части в пропорциях $\ell = 40\%$, $C = 40\%$ и $B = 20\%$. Выборки ℓ и C используются для вычисления критерия минимального смещения, а выборка B — для вычисления ошибки предсказания. Для критериев c_1 и c_2 выборка обычно делится на две равные части.

1.1.6. Парето-оптимальный фронт в пространстве критериев

Парето-оптимальный фронт — альтернатива комбинированным критериям. Выбирается множество внешних критериев, условиям оптимальности которых должна удовлетворять модель. Каждой модели ставится в соответствие вектор в пространстве выбранных критериев. Отыскиваются векторы, принадлежащие Парето-оптимальному фронту множества всех векторов, соответствующих порожденным моделям. При создании комбинированного критерия рассматриваются модели, критерии которых принадлежат полученному Парето-оптимальному фронту.

1.2. Алгоритм порождения моделей МГУА

Целю МГУА является получение модели в результате перебора моделей из индуктивно-порождаемого множества. Параметры каждой модели настраиваются так, чтобы доставить минимум выбранному внешнему критерию. Различают два основных типа алгоритмов МГУА — однорядный и многорядный.

Все алгоритмы МГУА воспроизводят схему массовой селекции: последовательно порождаются модели возрастающей сложности. Каждая модель настраивается — методом наименьших квадратов находятся значения параметров. Из моделей-претендентов выбираются лучшие в соответствии с выбранным критерием. Многорядные алгоритмы могут вычислять остатки регрессионных моделей после каждого ряда селекции или не вычислять; при этом используются исходные данные.

Каждая полиномиальная модель однозначно определяется набором индексов s входящих в нее мономов:

$$y = a_0 + \mathbf{w}(s) \cdot \mathbf{a}(s). \quad (1.2.1)$$

Элементы вектора $\mathbf{w}$ — коэффициенты при мономе полинома Колмогорова-Габора; элементы вектора $\mathbf{a}$ — результат произведения свободных переменных соответствующих мономов. Индексы $s \subseteq \{1, \ldots, F_0\}$ есть индексы мономов, входящих в модель. Иначе, произвольная модель:

$$y = w_0 + \mathbf{w}(s) \cdot \mathbf{a}(s) \quad (1.2.2)$$

порождается набором индексов $s \subseteq \{1, \ldots, F_0\}$, включающих соотвествующие элементы векторов

$$\mathbf{w} = \langle w_1, \ldots, w_{F_0} \rangle \quad \text{и} \quad \mathbf{a} = \langle a_1, \ldots, a_{F_0} \rangle.$$

При ограничении степени полинома числом R, число мономов полинома равно

$$F_0 = \sum_{r=1}^{R} \overline{C}_r^P = \sum_{r=1}^{R} \frac{(r+P-1)!}{P!(r-1)!}, \quad (1.2.3)$$

а число моделей первого ряда соответственно равно 2^{F_0}. Здесь $\overline{C}_r^P$ — число сочетаний с повторениями из P по r, P — число свободных переменных — элементов вектора x.

1.2.1. Комбинаторный алгоритм

Комбинаторный (однорядный) алгоритм использует только один ряд выбора. При этом порождаются все возможные линейные комбинации ограниченной сложности. Так как под сложностью понимается число линейно входящих параметров w, то сложность не превосходит заданное значение F_0. Пусть, как и ранее

$$y = w_0 + w_1 a_1 + w_2 a_2 + w_3 a_3, \ldots, w_{F_0} a_{F_0}. \quad (1.2.1.1)$$

Алгоритм выполняет следующие шаги. Для всех комбинаций входных аргументов строятся модели-претенденты неубывающей сложности. Например,

$$\begin{aligned}
y_1 &= w_{10} + w_{11} a_1, \\
y_2 &= w_{20} + w_{22} a_2, \\
y_3 &= w_{30} + w_{31} a_1 + w_{32} a_2, \\
y_4 &= w_{40} + w_{43} a_3, \\
y_5 &= w_{50} + w_{51} a_1 + w_{53} a_3, \\
y_6 &= w_{60} + w_{62} a_2 + w_{63} a_3, \\
y_7 &= w_{70} + w_{71} a_1 + w_{72} a_2 + w_{73} a_3.
\end{aligned}$$

Параметры каждой модели настраиваются методом наименьших квадратов по обучающей выборке. Наилучшая модель выбирается исходя из минимума значения внешнего критерия. Как вариант — назначается порог и выбираются несколько моделей, значения критерия для которых не превышает этот порог.

При программировании данного алгоритма удобно ввести переменную выбора — вектор $c(s) = \langle c_1, \ldots, c_{F_0} \rangle$. Его элемент $c_i \in \{0,1\}$ принимает значение 1, если $i \in s$, в противном случае 0. Тогда модель имеет вид

$$y = c(s)\cdot w\cdot a. \quad (1.2.1.2)$$

Последовательность векторов c для предыдущего примера выглядит как

$$\begin{matrix} 0 & \cdots & 0 & 1 \\ 0 & \cdots & 1 & 0 \\ 0 & \cdots & 1 & 1 \\ \vdots & \ddots & \vdots & \vdots \\ 1 & \cdots & 1 & 1 \end{matrix}.$$

Так как при порождении моделей необходимо выбирать из 2^{F_0} моделей, что может повлечь недопустимо большое время вычислений, предложены несколько эвристических алгоритмов, позволяющих сократить время вычислений, без уменьшения максимальной сложности моделей.

Многорядный алгоритм

На первом ряде алгоритма порождения моделей задано множество из F_0 переменных a_i. Порождаются модели как линейные комбинации всевозможных пар переменных

$$y_{(ij)} = w_0 + w_i a_i + w_j a_j, i,j = 1,\ldots,F_0, i \neq j. \quad (1.2.1.3)$$

Число моделей первого ряда M_1 есть число сочетаний $C_2^{F_0} = \frac{1}{2}F_0(F_0-1) = M_1$. Каждая модель, порождаемая на ряде задается парой индексов (i,j). После порождения моделей их параметры настраиваются с использованием внутреннего критерия. Затем выбираются F_1 наилучших моделей с использованием внешнего критерия. Эти модели используются в следующем ряде. Множество выбранных моделей задано множеством пар индексов $\{(i,j)\}$. На первом ряде индексы выбранных моделей i,j принадлежат множеству $s_1 \subset \{1,\ldots,F_0\}$.

На каждом последующем ряде новые модели — порождаются как суммы всевозможных пар выбранных моделей предыдущего ряда. Например, для второго ряда множество моделей

$$z_{(pquv)} = y_{(pq)} + y_{(uv)} = w_0 + w_p a_p + w_q a_q + w_u a_u + w_v a_v, p,q,u,v = 1,\ldots,s_1, p \neq q \neq u \neq v. \quad (1.2.1.4)$$

Порожденные модели снова настраиваются, выбираются наилучшие. Индексы, задающие выбранные модели принадлежат множеству $\{(p,q,u,v)\}$, где $p,q,u,v \in s_2 \subset s_1 \subset \{1,...,F_0\}$.

Таким образом на l-м ряде с помощью вышеприведенного алгоритма выбирается множество моделей, задаваемых множеством наборов индексов $\{p,...,v\}$, которые принадлежат полученному множеству $s_l \subset s_{l-1} \subset s_1 \subset \{1,...,F_0\}$. При этом индексация элементов моделей остается сквозной,

$$z_{(p...v)} = y_{(p...q)} + y_{(u...v)} = w_0 + w_p a_p + ... + w_v a_v. \quad (1.2.1.5)$$

Остановка порождения моделей на последующих рядах происходит в том случае, когда с увеличением номера слоя, то есть, с усложнением моделей, происходит увеличение внешнего критерия лучшей модели.

1.3. Начальный полином для МГУА

В качестве начального полинома будем использовать полином Колмогорова-Габора, имеющего вид:

(1.2.1.6)

где $g[x(t)]$– будущее предсказываемое значение функции; N – количество интервалов предыстории; x_{i_1}, x_{i_2} – значения функции в прошлом. Суммы в (1.2.1.6) охватывают всю предысторию и представляют своеобразную базу неявных знаний: первая сумма определяет обобщенный линейный предсказывающий фильтр; r_i– коэффициент, учитывающий влияние (вес) дискретного значения x_i стационарной функции, соответствующего моменту на i интервалов раньше, чем текущий момент; вторая сумма состоит из произведений этих дискретных значений по два, включая квадраты, а коэффициенты $r_{i_1 i_2}$ определяют веса пар их значений в моменты на i_1 и i_2 интервалов раньше текущего момента и т.д. Если предыстория предсказываемого процесса представлена m известными значениями, т.е. m – длина обучающей выборки, то, очевидно, оператор (1.2.1.6) содержит m членов 1-го порядка, $\frac{1}{2}m(m+1)$ членов 2-го порядка и т.д. Увеличение порядка оператора ведет к увеличению точности предсказания, но при этом количество членов оператора, а, следовательно, и объем вычислений резко возрастают, что привело к проблеме оптимизации соотношения количества членов оператора (1.2.1.6) и длины предыстории, а также практически приемлемого способа нахождения весовых коэффициентов r_i.

Известно, что весовые коэффициенты полинома Колмогорова-Габора (1.2.1.6) можно определить путем составления и решения системы нормальных уравнений Гаусса или при помощи методов адаптации, в частности стохастической аппроксимации.