Выбор переменных при идентификации многомерных систем

К. Р. Чернышев
Институт проблем управления им. В. А. Трапезникова Российской академии наук
may18_2005@mail.ru

***Аннотация*. В данной статье рассматривается проблема выбора входных и выходных переменных для системы с несколькими входами и выходами. В статье рассмотрен достаточно общий случай, когда выходные переменные модели представлены в виде условных математических ожиданий выходных переменных модели системы относительно (обобщенных) входных переменных. В статье показано, что правильный выбор переменных (как входных, так и выходных) осуществляется таким образом, чтобы применялись меры зависимости двумерных и многомерных случайных величин, состоятельных по Реньи (все определения даны в тексте). Также вводится неоднородность входных/выходных переменных и представляется обобщение аксиом Реньи на случай мер многомерной зависимости.**

Ключевые слова: идентификация систем, модель «вход-выход», состоятельная мера зависимости, мера неоднородности, условное математическое ожидание

I. Введение

Как известно, наилучшее приближение нелинейной зависимости, скажем y, является z условным математическим ожиданием $\mathbf{E}\{y/z\}$ [1]. То же самое справедливо и в отношении множественной зависимости, скажем вектора Z, где $Z=(z_1, \ldots, z_n)^T$, что означает $\mathbf{E}\{y/Z\}=\mathbf{E}\{y/z_1, \ldots, z_n\}$. Соответственно, к подобным задачам относятся многочисленные публикации, в частности, основанные на ядерной оценке регрессии [2–6].

При этом мало внимания уделяется выбору именно самих переменных, $Z=(z_1, \ldots, z_n)^T$, что означает в определенном смысле правильный выбор подмножества $Z_i=(z_{i_1}, \ldots, z_{i_m})^T$, $m \le n$,. С точки зрения теории идентификации переменные $Z=(z_1, \ldots, z_n)^T$ естественно называть входными переменными, а y – выходной переменной. То же самое справедливо и для подходящего выбора выходных переменных $Y=(y_1, \ldots, y_p)^T$, где $\mathbf{E}\{y_i/Z\}=\mathbf{E}\{y_i/z_1, \ldots, z_n\}$, $i=1, \ldots, p$. Это означает выбор подмножества $Y_j=(y_{i_1}, \ldots, y_{i_q})^T$, $q \le p$, подходящего в определенном смысле.

II. Краткий анализ мер зависимости

В общем случае решение задач вероятностной идентификации систем всегда основано на использовании мер зависимости случайных величин; в случае представления исследуемой системы с помощью отображений вход-выход и описания системы в пространстве состояний. Среди таких мер зависимости традиционная мера, основанная на линейной корреляции, является наиболее известной и широко используемой. Ее применимость непосредственно следует из того, что в основе задачи идентификации зачастую лежит среднеквадратический критерий. Основным преимуществом мер зависимости, основанных на линейной корреляции, является простота их использования, что связано с возможностью выявления требуемых свойств системы и относительной простотой построения оценок, в тои числе на основе наблюдения зависимых данных. Однако основным недостатком мер зависимости, основанных на линейной корреляции, является то, что они могут обращаться в ноль даже при наличии детерминированной зависимости между случайными величинами. Например, в случае квадратичной зависимости, $y(t)=z^2(t)$, где $z(t)$ является стационарным гауссовским процессом с нулевым средним и единичной дисперсией [7], или нечетной зависимостью, скажем $y(t)=5z^3(t)-3z(t)$, с маргинальным распределением $z(t)$, которое является равномерным на интервале $(-1,+1)$ [8].

Кроме того, это относится к случаю стохастической зависимости, скажем, в отношении распределения вероятностей класса $p_{zy;\lambda}(z,y)$ О.В. Сарманова [9, 10]

$$p_{zy;\lambda}(z,y) = p_z(z)p_y(y)\left(1+\lambda\phi_1(z)\phi_2(y)\right), \tag{1a}$$

где

$$\int p_z(z)\phi_1(z)dz=0, \tag{1b}$$

$$\int p_y(y)\phi_2(y)dy=0, \tag{1c}$$

$$\lambda\phi_1(z)\phi_2(y)\ge -1. \tag{2}$$

В формулах (1), $p_z(z)$, $p_y(y)$ – маргинальные плотности распределения вероятностей, $\phi_1(z)$, $\phi_2(y)$ – функции, удовлетворяющие условию (2) совместно со скаляром λ.

Коэффициенты корреляции и корреляционные отношения часто используются в качестве мер зависимости, основанных на сравнении характерных моментов совместного и маргинального распределений вероятностей пары представляющих интерес случайных величин.

$$\theta(z,y)=\frac{\mathbf{var}\left(\mathbf{E}\left\{ y/z \right\}\right)}{\mathbf{var}(y)}, \quad \mathbf{var}(y)>0,$$

и максимальный коэффициент корреляции

$$R(z,y)=\sup_{\{B\},\{C\}} \mathbf{corr}(B(y),C(z)),$$

$$\mathbf{var}(B(y))>0, \quad \mathbf{var}(C(z))>0.$$

В данных формулах $\mathbf{var}(\cdot)$ обозначает дисперсию. Супремум в максимальном коэффициенте корреляции берется по множествам измеримых по Борелю функций $\{B\}$ и $\{C\}$, а также , $B\in\{B\}$, $C\in\{C\}$. Из них только максимальный коэффициент корреляции удовлетворяет всем аксиомам для мер зависимости [8], в то время как обычная корреляция и корреляционное отношение – нет. Меры зависимости, которые обращаются в ноль лишь при зависимости случайных величин (векторов), следует называть состоятельными, или состоятельными по Колмогороову.

Наряду с мерами зависимости, основанными на корреляции широкий класс мер зависимости строится прямым сравнением совместного $p_{yz}(y,z)$ и маргинальных, $p_y(y)$, $p_z(z)$ вероятностных распределений случайных величин. Такой класс называется мерами дивергенции вероятностных распределений. Наиболее .известными представителями класса являются дифференциальные взаимные информационные меры,

$$0\le I(z,y)=\mathbf{E}\left\{\ln\frac{p_{zy}(z,y)}{p_z(z)p_y(y)}\right\}\le\infty,$$

и коэффициент сопряженности

$$0\le\Delta^2(z,y)=\mathbf{E}\left\{\frac{(p_{zy}(z,y)-p_z(z)p_y(y))^2}{p_{zy}(z,y)p_z(z)p_y(y)}\right\}\le\infty.$$

Здесь $\mathbf{E}\{\cdot\}$ обозначает математическое ожидание. Эти меры зависимости состоятельны по Колмогорову, но в то же время они не удовлетворяют аксиомам Реньи C, E и G.

Те меры зависимости, которые удовлетворяют всем аксиомам Реньи, кроме, быть может, аксиомы F, а именно инвариантности к взаимно однозначным преобразованиям случайных величин, которые мы будем называть состоятельными по Реньи.

III. Снижение взаимной зависимости входных переменных

В практических задачах возможны взаимные влияния входных переменных друг на друга, причем такие взаимоотношения могут быть столь существенными, что одна входная переменная фактически является детерминированной функцией другой входной переменной. Такие взаимосвязи естественным образом отражаются в значении показателя, основанном на состоятельных по Реньи мерах зависимости, представляющий собой взаимосвязь между выходными и входными переменными. Поэтому предлагается следующий алгоритм для надлежащего исключения входных переменных из модели. Для входной переменной с наибольшей состоятельной по Реньи мерой зависимости с выходной переменной

$$\mu_R(y,z_{i_0})\ge\mu_R(y,z_{i_j}) \quad \forall j\ne 0$$

рассчитываются состоятельные по Реньи меры зависимости с каждой из других *входных* переменных,

$$\mu_R(z_{i_0},z_{i_j}) \quad \forall j\ne 0\,.$$

Если значения этих показателей зависимости больше, чем значение κ, указанное исследователем,

$$\mu_R(z_{i_0},z_{i_k})\ge\kappa,$$

то данная переменная исключается из конструируемой модели. Это связано с тем, что вклад в выходную переменную вносила входная переменная, которая относительно имеет наибольшее значение меры зависимости с выходной переменной.

IV. Распространение аксиом Реньи на случай многомерной зависимости

Подход к построению состоятельных по Реньи мер зависимости позволяет детально распространить аксиомы Реньи на случай многомерной зависимости и, следовательно, состоятельные по Реньи меры зависимости для случайных векторов. В литературе расширение аксиом Реньи на многомерный случай было представлено в работе [11]. Однако оно не обеспечивает точного согласия с аксиомами Реньи для случая двумерного распределения, удовлетворение которому было бы естественно.

В частности, это касается аксиом C, E и G, связанных с условиями нормировки. С другой стороны, условие нормировки очень важно, так как принятие значений на всей положительной полуоси фактически не дает основы для сравнения и оценки значимости тех или иных величин. Аксиомы, предложенные в этой статье, при подходящем сравнении с аксиомами [11], заключаются в следующем.

A) $\mu(z_1,\dots,z_n)$ определяется для любого случайного вектора $Z=(z_1,\dots,z_n)$, ни один из компонентов вектора не Z является константой с вероятностью 1.

B) При любой перестановке $\sigma=(i_1,\dots,i_n)$ индексов $\{1, 2, \dots, n\}$ мера инвариантна, т.е.

$$\mu(Z)=\mu(z_1,\dots,z_n)=\mu(z_{i_1},\dots,z_{i_n}).$$

C) $0\le\mu(z_1,\dots,z_n)\le 1$ (против $0\le\mu(z_1,\dots,z_n)=\gamma$ (может быть ∞) в [11]).

D) $\mu(z_1,\dots,z_n)=0$ тогда и только тогда, когда случайные величины $z_1,\dots,z_n$ независимы.

E) $\mu(z_1,\ldots,z_n)=1$ против $0 \le \mu(z_1,\ldots,z_n)=\gamma$ (может быть ∞ в [11]) тогда и только тогда, когда существует детерминированная зависимость между компонентами случайного вектора Z .

F) Инвариантность к любому взаимно однозначному $\Psi=(\psi_1,\ldots,\psi_n)$ преобразованию $Z=(z_1,\ldots,z_n)$ на R^n, т.е. $\Psi(Z)=(\psi_1(z_1),\ldots,\psi_n(z_n))$, а именно:

$$\mu(\psi_1(z_1),\ldots,\psi_n(z_n))=\mu(z_1,\ldots,z_n).$$

G) Если совместное распределение вектора $Z=(z_1,\ldots,z_n)$ является нормальным, то для случая $n=2$ по $\mu(z_1,z_2)=|\mathbf{corr}(z_1,z_2)|$ (в [11] $\mu(z_1,z_2)$ является возрастающей функцией $\mathbf{corr}(z_1,z_2)$).

V. Выбор выходных переменных

Состоятельная по Реньи мера множественной зависимости может быть использована для адекватного построения неоднородности выходных переменных исследуемой системы и выявления наиболее значимых из них. То есть можно предположить, что выходные переменные системы отсортированы в порядке возрастания значений состоятельной по Реньи меры множественной зависимости для каждого выхода и всех входных переменных,

$$\begin{aligned}\mu_1=\mu_R(y_{i_1},z_1,\ldots,z_{p_1})\le\ldots\le\\ \le\mu_m=\mu_R(y_{i_M},z_1,\ldots,z_{p_l})\end{aligned} \qquad (3)$$

В (3) нижние буквы $1,\ldots,p_1,\ldots,\quad 1,\ldots,p_l$, используются для обозначения входных переменных, которые были выбраны в соответствии с описанной выше процедурой выбора входных переменных.

Затем неоднородность эффектов всех входных переменных (выбранных) на каждую выходную переменную целевой системы может быть выражена количественно с помощью следующих мер, которые принимают значения в единицах интервалов.

$$\eta^{out}=\frac{\sum_{k=1}^{p}\nu_k\mu_k}{\sum_{k=1}^{p}\mu_k}, \qquad (4)$$

где

$$\nu_k=2\frac{k-1}{p-1}-1.$$

Характеристики меры неоднородности выходной переменной такие же, как и у входных переменных.

Принимает значения в единичном интервале,

$$0\le\eta^{out}\le 1.$$

Она становится равной нулю, только если значение состоятельной по Реньи меры множественной зависимости между всеми выходными и входными переменными совпадает.

$$\eta^{out}=0 \text{ тогда и только тогда, когда } \mu_1=\ldots=\mu_p.$$

Обращение в единицу достигается только в том случае, если мера множественной зависимости между выходом и всеми входными переменными не равна нулю, ее значение согласуется с концепцией Реньи для всех остальных выходных переменных, а мера множественной зависимости равна нулю,

$$\max_{\mu_1,\ldots,\mu_m}\{\eta^{out}\}=1,$$

$$\arg\max_{\mu_{i_1},\ldots,\mu_{i_p}}\{\eta^{out}\}=(0,\ldots,0,\mu)\quad\forall\mu>0.$$

Соответственно решение задачи включения или исключения той или иной выходной переменной решается следующим образом. Исследователь задает минимально допустимую величину состоятельной по Реньи меры множественной зависимости $\mu_{\min}$и максимально допустимую величину меры неоднородности $\eta^{\max}$. Эти величины образуют прямоугольник

$$\Delta=[0,\mu_{\min}]\otimes[\eta^{\max},1] \qquad (5)$$

и если выходная переменная попадает в прямоугольник (5), она исключается из модели.

VI. Заключение

Статья посвящена наиболее общему случаю выбора переменных в модели нелинейной системы, представленной в виде условного математического ожидания. При этом во входную переменную $Z=(z_1,\ldots,z_n)^T$ могут входить как входные переменные системы, так и предшествующие выходные переменные, и в статье рассмотрены проблемы выбора этих переменных (их количества и качества).

В основе всего подхода лежит последовательное применение концепции характеристик зависимости по Реньи и характеристик неоднородности, которые позволяют выбирать входные и выходные переменные. И, конечно же, на первом плане остается роль исследователя, соответствующим образом выбирающего минимально допустимый размер состоятельной по Реньи меры зависимости и максимально допустимый размер меры неоднородности.

Предложенные подход является независимым от распределения случайных величин/векторов, а, с другой стороны, ориентирован на *любую* состоятельную по Реньи меру зависимости.

Список литературы

[1] Watson G.S. Smooth regression analysis // Sankhya. Ser. A. 1964. Vol. 26, P. 355-372.

[2] Васильев В.А., Добровидов А.В., Кшкин Г.М. Непараметрическое оценивание функционалов от распределений от стационарных последовательностей. М.: Наука, 2004. 508 с.

[3] Fahrmeir L., Kneib T., Lang S., Marx B.D. Regression. Models, Methods and Applications. Springer, 2021. 746 p.

[4] Györfi L., Kohler M., Krzyżak A., Walk H. A Distribution-Free Theory of Nonparametric Regression. Springer, 2002, 650 p.

[5] Nelles O. Nonlinear System Identification. From Classical Approaches to Neural Networks, Fuzzy Models, and Gaussian Processes. Springer, 2020, 1226 p.

[6] Pillonetto G., Tianshi Chen, Chiuso A., De Nicolao G., Ljung L. Regularized System Identification. Learning Dynamic Models from Data. Springer, 2022, 394 p.

[7] Rajbman N.S. Extensions to nonlinear and minimax approaches // Trends and Progress in System Identification. Ed. P. Eykhoff. Elsevier, 1981. Chapter 7. P. 185-237.

[8] Rényi A. On measures of dependence // Acta Mathematica Academiae Scientiarum Hungarica. 1959. Vol. 10, No. 3-4. P. 441-451.

[9] Сарманов О.В. Замечания о некоррелированных гауссовских зависимых случайных величинах, Теория вероятностей и ее применения. 1967. Т. 12:№ 1. С. 141-143.

[10] Balakrishnan N., Chin-Diew Lai. Continuous Bivariate Distributions. Second Edition. Springer, 2009, 712 p.

[11] Micheas A.C., Zografos K. Measuring stochastic dependence using φ-divergence // J. Multivariate Analysis. 2006. Vol. 97. P. 765-784.