УДК 519.71

МЕТОДЫ ИДЕНТИФИКАЦИИ СИСТЕМ ПО КРИТЕРИЯМ МИНИМУМА ЭНТРОПИИ ОШИБКИ

К.Р. Чернышев
Институт проблем управления им. В.А. Трапезникова РАН
Россия, 117997, Москва, Профсоюзная ул., 65
E-mail: myau@ipu.ru

Ключевые слова: энтропия Реньи, наихудшая плотность распределения вероятностей, идентификация системы, минимаксный подход, выбор входных переменных, вектор параметров.

Аннотация. Рассмотрена постановка задачи идентификации параметров входо/выходного отображения систем, основанная на квадратичной энтропии в смысле Реньи ошибки идентификации. При этом критерий идентификации максимизируется по всем возможным распределениям с последующей минимизацией по вектору параметров системы. Кроме того, квадратичная энтропия Реньи используется для построения меры неравномерности, позволяющей адекватным образом, обоснованным свойствами предлагаемой меры неравномерности, выбирать входные переменные (идентификация структуры как предварительный шаг) модели, подлежащей далее параметрической идентификации. При этом особое значение имеет свойство нормированности состоятельной меры зависимости, которой и обладает предложенная мера.

Сложность и многообразие процессов, протекающих в нелинейных динамических системах, не позволяют создать какой-либо универсальный алгоритм идентификации таких систем, что, обычно, приводит к необходимости привлечения априорных предположений об их структуре. Например, это связано с отнесением структуры моделей систем к классу блочно-ориентированных моделей – Винера, Гаммерштейна, различных вариаций на их основе. Из-за системного характера идентификации в процессе идентификации возникает системный парадокс, однако до сих пор не разработан научно обоснованный метод выхода из этого парадокса, что, несомненно, свидетельствует о трудности решения проблемы идентификации нелинейных систем.

Традиционно неопределенность количественного показателя систем описывается его энтропией, а энтропия Шеннона (в случае дискретных распределений) и дифференциальная энтропия (в случае непрерывных распределений) является наиболее известной и используемой в этом контексте. В то же время, наряду с определением энтропии Шеннона (дифференциальной энтропии), известны и другие подходы к определению энтропии случайной величины, при этом в таких определениях энтропия Шеннона используется как частный случай. Так для случайного вектора X, имеющего соответствующую многомерную плотность распределения вероятностей $p_X(X)$, энтропия Реньи порядка α [1] определяется как

(1) $$H_\alpha^R(p_X)=\frac{1}{1-\alpha}\ln\left(\mathbf{E}\left(p_X(X)\right)^{\alpha-1}\right), \alpha>0, \alpha\neq 1.$$

В (1) $\mathbf{E}(\cdot)$ – математическое ожидание. Одновременно в рамках определения (1) при стремлении α к 1 в силу правила Лопиталя $H_\alpha^R(p_X)$ стремится к выражению, определяющему энтропию дифференциальную энтропию, которую можно рассматривать как энтропию Реньи первого порядка.

Кроме того, энтропия Реньи была общепризнанна как имеющая более привлекательную конструкцию по сравнению с дифференциальной энтропией, поскольку первая представляет собой «логарифм интеграла», т.е. что вычислительно проще, чем «интеграл от логарифма» в случае дифференциальной энтропии. Между тем выбор самого порядка α играет важную роль, поскольку, чем больше порядок, тем сложнее становится соответствующая вычислительная процедура. Среди различных значений порядка $\alpha > 0, \alpha \neq 1$ все могут быть применены для расчета энтропии, но сложность задачи будет расти экспоненциально по α, в то время как в литературе общепризнанна целесообразность применения значения α, равного 2 [2]. Для $\alpha = 2$ величина

(2) $$H_2^R(p_X) = -\ln\left(\mathbf{E}(p_X(X))\right)$$

называется квадратичной энтропией Реньи.

В настоящей работе в энтропийных терминах рассматривается задача идентификации системы. А именно, в качестве критерия идентификации, подлежащего оптимизации, рассматривается квадратичная энтропия Реньи ошибки идентификации.

Рассмотрим задачу идентификации параметров нелинейной стохастической системы с дискретным временем, описываемой входо/выходным отображением, по наблюдениям ее одномерного (скалярного) выходного процесса $y(t)$, $t = 1, 2, \dots$ и входного n_U-мерного процесса $U(t) = \left(u_1(t), \dots, u_{n_U}(t)\right)^T$, $t = 1, 2, \dots$. Входные и выходные процессы систем предполагаются стационарными и стационарно-связанными в узком смысле. Модель системы ищется в виде

(3) $$\hat{y}(t;\Theta) = \psi\left(\Theta;\ y(t-1), \dots, y(t-q_y), U(t), \dots, U(t-q_U)\right),$$

где $\hat{y}(t;\Theta)$ – выходной процесс модели системы, $\psi(\cdot;\cdot,\dots,\cdot,\cdot,\dots,\cdot)\colon R^{n_\Theta + q_y + n_U \times (q_U+1)} \to R^1$ – известная функция, $\Theta = \left(\theta_1, \dots, \theta_{n_\Theta}\right)^T$ – n –мерный вектор неизвестных параметров системы, q_y и q_U – заданные порядки системы. Вектор неизвестных параметров системы $\Theta = \left(\theta_1, \dots, \theta_{n_\Theta}\right)^T$ подлежит идентификации по минимаксному критерию

(4) $$\min_\Theta \left\{ \max_{\mathbf{P}_{cont}|_{\mathbf{var}(\varepsilon(t;\Theta))}} H_2^R\left(p_{\varepsilon(t;\Theta)}\right) \right\},$$

где $\varepsilon(t;\Theta)$ – ошибка идентификации, $H_2^R\left(p_{\varepsilon(t;\Theta)}\right)$ – квадратичная энтропия Реньи (2) непрерывной кумулятивной случайной величины $\varepsilon(t;\Theta)$ (5),

(5) $$\varepsilon(t;\Theta) = y(t) - \psi\left(\Theta;\ y(t-1), \dots, y(t-q_y), U(t), \dots, U(t-q_U)\right),$$

где $\mathbf{var}(\cdot)$ – дисперсия случайной величины, $\mathbf{P}_{cont}|_{\mathbf{var}(\cdot)}$ – класс *всех* непрерывных одномерных вероятностных распределений с дисперсией $\mathbf{var}(\cdot)$. Между тем, поскольку энтропия случайной величины не зависит (за некоторым исключением) от ее математического ожидания, критерий (4) естественно дополнить условием совпадения математических ожиданий выходных переменных системы и модели:

(6) $$\mathbf{E}(y(t)) = \mathbf{E}(\hat{y}(t;\Theta)).$$

В первую очередь следует отметить, что при заданной дисперсии непрерывной случайной величины максимальная величина квадратичной энтропии Реньи (выражение в фигурных скобках в формуле (4)) строится в соответствии с результатами, представленными в статьях [3-5]. В конечном итоге на основе [3-5] можно получить

(7) $$g_{\alpha,C_X}(X) = A_{\alpha,n_X}\left(\left(1-(\alpha-1)b_{\alpha,n_X}X^T[C(X)]^{-1}X\right)_+\right)^{1/(\alpha-1)},$$

(8) $$b_{\alpha,n_X} = \frac{1}{2\alpha - n_X(1-\alpha)},$$

(9) $$A_{\alpha,n_X} = \frac{\Gamma\left(\frac{\alpha}{\alpha-1}+\frac{n_X}{2}\right)\left(b_{\alpha,n_X}(\alpha-1)\right)^{n_X/2}}{\Gamma\left(\frac{\alpha}{\alpha-1}\right)\pi^{n_X/2}\sqrt{\det(C(X))}},$$

и для любой скалярное величины χ

(10) $$\chi_+ = \max\{\chi, 0\}.$$

Тогда на основании (7)-(10) для квадратичной энтропии $H_2^R(x)$ Реньи одномерной случайной величины x с дисперсией достигается $\mathbf{var}(x)$ максимум $H_2^R(x)$ при следующей плотности распределения вероятностей:

(11) $$g_{2,\mathbf{var}(x)}(x) = \frac{3}{4\sqrt{5\mathbf{var}(x)}}\left(1-\frac{x^2}{5\mathbf{var}(x)}\right)_+.$$

Соответственно, поскольку ошибка идентификации (5) имеет плотность распределения вероятности (11), выражение в фигурных скобках в формуле (4) принимает вид

(12) $$\max_{\mathbf{P}_{cont}|_{\mathbf{var}(\varepsilon(t;\Theta))}} H_2^R(\varepsilon(t;\Theta)) = \ln\left(\frac{5\sqrt{5}}{3}\right) + \frac{1}{2}\ln\left(\mathbf{var}(\varepsilon(t;\Theta))\right).$$

Таким образом, в силу формулы (12) и совпадения экстремумов функции и ее логарифма, из критерия (4)-(6) следует следующая система уравнений для определения вектора Θ неизвестных параметров:

(13) $$\mathrm{C}(\varepsilon(t;\Theta), \nabla_\Theta\varepsilon(t;\Theta)) = \mathbf{0},$$

где $\mathrm{C}(\cdot,\cdot)$ есть вектор-столбец (в данном случае n_Θ-мерный), компоненты которого являются ковариациями случайного значения (в данном случае это случайное значение $\varepsilon(t;\Theta)$) и соответствующей компоненты случайного вектора (в данном случае этого вектора $\nabla_\Theta\varepsilon(t;\Theta)$), $\mathbf{0}$ является n_Θ-мерным нулевым вектором-столбцом.

В дополнение к уравнению (13) одновременно утверждается, что оценки, полученные с помощью обычного среднеквадратичного критерия $\min_\Theta \mathbf{var}(\varepsilon(t;\Theta))$, обладает минимаксным свойством ошибки идентификации, понимаемой как разность выходных переменных системы и модели в смысле квадратичного энтропийного критерия Реньи, выраженного формулами (4)-(6).

Модель (3) может быть представлена в виде $\hat{y}(t;\Theta) = \tilde{\psi}(\Theta; Z(t))$ или же

(14) $$\hat{y}(t;\Theta) = \tilde{\psi}(\Theta; z_1(t), \dots, z_{n_Z}(t))$$

с обобщенным входным вектором $Z(t) = \left(z_1(t), \dots, z_{n_Z}(t)\right)^T$, состоящим из соответствующих предшествующих выходных и входных переменных; таким образом, размерность обобщенного входного вектора $Z(t)$ равна $n_Z = q_y + n_U \times (q_U + 1)$. Применительно к унифицированному представлению (14) выбор конкретной входной переменной $z_i(t)$, $i = 1, \dots, n_Z$ является отдельной специфической проблемой, связанной с идентифицируемостью системы и идентификацией структуры.

Естественным выводом для принятия адекватного решения о выборе входных переменных является учет неоднородности влияния входных переменных на выходную. Для описания неоднородности можно применить подход [7], который основан на мерах зависимости, удовлетворяющих аксиомам [6] (за исключением, возможно, аксиомы F). Такие меры будем в дальнейшем называть состоятельными по Реньи и обозначать буквой «*R*» в качестве верхнего индекса, т. е. $\mu^R(\xi,\zeta)$.

Пусть входные переменные $x_1, \dots, x_N$ системы упорядочены как $x_{i_1}, \dots, x_{i_N}$ в соответствии с возрастающими значениями состоятельной меры зависимости между выходной и каждой из входных переменных:

(15) $$\mu_1 = \mu^R(y, x_{i_1}) \le \dots \le \mu_N = \mu^R(y, x_{i_N}).$$

Тогда, с учетом (15), такая неоднородность может быть *количественно* выражена следующей мерой [7]

(16) $$\eta_h = \frac{\sum_{k=1}^{N}\left(2\frac{k-1}{N-1}-1\right)\mu_k}{\sum_{k=1}^{N}\mu_k}.$$

Выражение (16) принимает свои значения на единичном интервале. При этом

1) $0 \le \eta_h \le 1$;

2) $\eta_h = 0$ тогда и только тогда, когда $\mu_1 = \dots = \mu_N$;

3) $\max\limits_{\mu_1,\dots,\mu_N} \{\eta_h\} = 1$;

4) $\operatorname{argmax}\limits_{\mu_{i_1},\dots,\mu_{i_N}}\{\eta_h\} = \left(\underbrace{0, \dots, 0}_{N-1}, \mu\right) \forall \mu > 0$.

При таком подходе каждая входная переменная x_i, $i = 1.\dots, N$ характеризуется двумя числами, каждое из которых принимает свои значения в единичном интервале. Первая – это значение состоятельной по Реньи меры зависимости $\mu^R(y, x_i)$, $i = 1.\dots, N$, вторая – мера неоднородности (16), общая для всех входных переменных. Следовательно, каждая входная переменная x_i может $i = 1.\dots, N$ быть связана с точкой $\mathrm{X}_i = (\mu^R(y, x_i);\, \eta_h)$ на единичном квадрате. В свою очередь, прерогатива исследователя состоит в том, чтобы задать минимально допустимую общую величину для всех $\mu^R(y, x_i)$, $i = 1.\dots, N$ скажем, $\underline{\mu^R}$, и максимально допустимую величину меры неоднородности в (16), скажем, $\overline{\eta_h}$. Соответственно, эти два значения образуют прямоугольник в единичном квадрате,

(17) $$U_{\mu\eta} = \left[0;\ \underline{\mu^R}\right] \times \left[\overline{\eta_h};\ 1\right]$$

Следовательно, если $\mathrm{X}_i \in U_{\mu\eta}$ в (17), то переменная x_i подлежит исключению из построенной модели, если она не принадлежит, $\mathrm{X}_i \notin U_{\mu\eta}$, то переменная подлежит вовлечению в модель. При этом исключаемая переменная заменяется на параметр. В свою очередь, если параметры в модели образуют функцию только от самих параметров, не включая переменные, то такая функция сама рассматривается (и идентифицируется) как один параметр.

Поскольку аксиомы Реньи предъявляют достаточно жесткие требования к мерам зависимости, класс состоятельных по Реньи мер зависимости, очевидно, значительно уже, чем класс «просто» согласованных мер зависимости. А именно, можно в конечном счете, получить

(18) $$d_2^R\left(p_{yz}(y,z)\,\middle\|\, p_y(y)p_z(z)\right) = 2 \times \sqrt{1 - \frac{2-\sqrt{4-3e^{-4D_2^R\left(p_{yz}(y,z)\| p_y(y)p_z(z)\right)}}}{e^{-4D_2^R\left(p_{yz}(y,z)\| p_y(y)p_z(z)\right)}}}.$$

Мера зависимости $d_2^R\left(p_{yz}(y,z)\,\middle\|\, p_y(y)p_z(z)\right)$ в (18) удовлетворяет всем аксиомам Реньи за исключением аксиомы F, а аксиоме F удовлетворяет мера (18) для аффинных преобразований случайных величин. В любом случае, $d_2^R\left(p_{yz}(y,z)\,\middle\|\, p_y(y)p_z(z)\right)$ в (14) может быть успешно использован в качестве состоятельной по Реньи меры зависимости для вычисления меры неоднородности системы, подлежащей идентификации, при выборе входных переменных.

Существует множество примеров, когда применение традиционных корреляционных методов при моделировании систем не дает подходящих результатов. Среди таких систем можно выделить те, для которых зависимость между входными и выходными переменными описывается плотностями распределения вероятностей, принадлежащими классу двумерных вероятностных распределений О.В. Сарманова [8].

Модель (6) – это результат последовательного выполнения следующих операций. На исходном этапе для заданной исследователем модели с изначально заданными входными переменными и параметрами, производится выбор (обобщенных) входных переменных в соответствии с представленным алгоритмом и на основе применения состоятельной меры зависимости (18). При этом, исключаемая из исходно заданной аналитической модели входная переменная заменяется на константу, подлежащую идентификации. С другой стороны учитывается, что любые функции от констант есть, в свою очередь, тоже константа, которая заменяет собой всю эту функцию из констант. Таким образом и формируется вектор параметров модели (3) Θ. То есть, это этап структурной идентификации. На следующим за ним, этапе параметрической идентификации, окончательно сформированный вектор Θ подлежит идентификации в соответствии с вышеописанной процедурой по минимаксному критерию на основе квадратической энтропии Реньи.

Список литературы

1. Rényi A. On measures of information and entropy // Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability. Berkley, 1960(1961). P. 547-561.
2. Principe J.C. Information theoretic learning. New York, Dordrecht, Heidelberg, London: Springer, 2010. 516 p.
3. Costa J., Hero A., Vignat C. On Solutions to Multivariate Maximum $\alpha-$Entropy Problems, Energy Minimization Methods in Computer Vision and Pattern Recognition // Springer, 2003. Vol. 2683 of the series Lecture Notes in Computer Science. P. 211-226.
4. Johnson O., Vignat C. Some results concerning maximum Rényi entropy distributions // Annales de l'Institut Henri Poincare. (B) Probability and Statistics. 2007. Vol. 43, No. 3. P. 339-351.
5. Tanaka H.-A., Nakagawa M., Oohama Y. A Direct Link between Rényi–Tsallis Entropy and Hölder's Inequality—Yet Another Proof of Rényi–Tsallis Entropy Maximization // Entropy. 2019. Vol. 21. 549.
6. Rényi A. On measures of dependence // Acta Math. Hung. 1959. Vol. 10, No. 3-4. P. 441-451.
7. Chernyshov K.R. System Identifiability and Structure Identification: Input and Output Variables Selection Based on Consistent Measures of Dependence // IFAC-PapersOnline. 2021. Vol. 54, No. 14. P. 132-137.
8. Sarmanov O.V. Remarks on uncorrelated Gaussian dependent random variables // Theory Probab. Appl. 1967. Vol. 12, No. 1. P. 124-126.