УДК 007 : 681.3.06

ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ ДИСКРЕТИЗАЦИИ ОБЛАСТЕЙ ИХ ЗНАЧЕНИЙ

В. Я. Пивкин[1], **И. В. Пивкина**[2]

[1] *Учреждение Российской академии наук*
Институт автоматики и электрометрии Сибирского отделения РАН,
630090, г. Новосибирск, просп. Академика Коптюга, 1
E-mail: piv@idisys.iae.nsk.su
[2] *New Mexico State University, Department of Computer Science,*
P.O. Box 30001, MSC CS, Las Cruces, NM 88003

Предложен способ обработки данных наблюдений, основанный на их дискретизации и графическом анализе. Цель обработки — определение и моделирование функциональной зависимости между входами и выходом экспериментального объекта. Рассмотрен случай серии экспериментов.

Ключевые слова: экспериментальные данные, функциональная зависимость, дискретизация, моделирование, серия экспериментов.

Исходные данные. Рассматриваются объекты или процессы, предположительно обладающие определённой, но неизвестной функциональной зависимостью между независимыми величинами (входами) и зависимой величиной (выходом). Наблюдаемый процесс характеризуется набором $\Omega = \{Y, X_1, \ldots, X_k\}$ параметров (Y — выход, $X = \{X_1, \ldots, X_k\}$ — входы), измеряемых в дискретные такты времени. Результаты эксперимента представлены совокупностью временных рядов $\Psi(\Omega) = \{y(j), x_1(j), \ldots, x_k(j)\}$ значений выхода и входов, наблюдаемых в моменты времени j, $j = 1, \ldots, N$.

Предполагается, что значение $y(j)$ выхода Y на такте времени j зависит от значений входа $x(j) = (x_1(j), \ldots, x_k(j))$, а области значений параметров — отрезки $I(Y), I(X_1), \ldots, I(X_k)$ действительной оси.

Вспомогательная информация включает в себя максимальные и минимальные значения параметров (области определения) и наборы их дискретизаций.

Представленная работа — это модернизация способа дискретизации экспериментальных данных, предложенного в [1, 2] для поиска функциональных зависимостей и построения реализующих их нечётких моделей (здесь вместо нечётких множеств используются обычные множества).

Анализ данных в случае одного эксперимента. Производится дискретизация данных, заключающаяся в разбиении областей их значений на непересекающиеся полуоткрытые и замкнутые интервалы и замене значений экспериментальных данных номерами содержащих их интервалов. В результате область значений входов разбивается на непересекающиеся подобласти, кодируемые совокупностью k-мерных целочисленных векторов, а область значений выхода представляется множеством непересекающихся интервалов U, кодируемых их номерами. Дискретизацию параметра будем называть дискретизацией размера n, если она заключается в разбиении области значений этого параметра на n интервалов, и равномерной в случае разбиения на интервалы равной длины. В предлагаемой работе дискретизации всех параметров равномерные.

Дискретизация выхода Y — равномерная размера $n(y)$. Границы значений выхода в экспериментальных данных равны $\min\limits_{y \in I(Y)} y$ и $\max\limits_{y \in I(Y)} y$. Длина элемента дискретизации выхода

$$\Delta(y) = \Big(\max_{y\in I(Y)} y - \min_{y\in I(Y)} y\Big)\Big/ n(y).$$

Последовательность кодов элементов дискретизации — возрастающая от 1 до $n(y)$. Дискретизированный выход со значением u $(1 \le u \le n(y))$ соответствует значениям экспериментальных данных от $\min\limits_{y\in I(Y)} y + (u-1)\Delta(y)$ до $\min\limits_{y\in I(Y)} y + u\Delta(y)$.

Дискретизация входа X_i, $i = 1, \ldots, k$, — равномерная размера $n(x_i)$. Границы значений входа в экспериментальных данных равны $\min\limits_{x_i(j)\in I(X_j)} x_i(j)$ и $\max\limits_{x_i(j)\in I(X_j)} x_i(j)$. Длина элемента дискретизации входа

$$\Delta(x_i) = \Big(\max_{x_i(j)\in I(X_j)} x_i(j) - \min_{x_i(j)\in I(X_j)} x_i(j)\Big)\Big/ n(x_i).$$

Последовательность кодов элементов дискретизации — возрастающая от 1 до $n(x_i)$.

Дискретизация преобразует $\Psi(\Omega)$ в совокупность целочисленных временных рядов $\Psi(D(\Omega))$. Данные j-го такта наблюдений образуют в $\Psi(D(\Omega))$ вектор $S(j) = (u(j), v(j))$, который состоит из выхода $u(j)$ и вектора входа $v(j) = (v^1(j), \ldots, v^k(j))$. Пусть $V_{\text{м}} = \{v_1, \ldots, v_p\}$ — множество неодинаковых (попарно не совпадающих) векторов входов из $\Psi(D(\Omega))$. Каждому v_i, $i = 1, \ldots, p$, из $V_{\text{м}}$ соответствует подпоследовательность U_i значений выхода с указанием числа вхождений s_k в U_i каждого из выходов u_k, $k = 1, \ldots, n(y)$. Будем считать, что $s_k = 0$, если $u_k \notin U_i$. Каждая подпоследовательность U_i исследуется как совокупность независимых измерений величины u_i. Условием существования функциональной зависимости является концентрация значений выхода в каждой из U_i в окрестности некоторого среднего (центра тяжести). Это среднее значение становится значением функции в случае, когда функциональная зависимость имеется. Дискретные данные при наличии функциональной зависимости являются базой (основой) для построения дискретной модели объекта или процесса.

В случае одного эксперимента для каждого вектора входа v_i значения выхода U_i упорядочиваются от его минимального значения $\min\limits_{u\in U_i} u$ до максимального $\max\limits_{u\in U_i} u$. Эти значения задают интервал изменения дискретизированного выхода u для данного входа v_i. Найдём координату центра тяжести дискретизированного выхода для входа v_i:

$$u(U_i) = \sum_{s_k\neq 0} u_k s_k \Big/ \sum_{s_k\neq 0} s_k.$$

Таким образом, для дискретной модели определён набор характеризующих параметров: размер модели и центры тяжести выходов U_i, $i = 1, \ldots, p$.

Во многих случаях возникает необходимость повторной дискретизации в целях улучшения параметров построенной дискретной модели. Основной операцией повторной дискретизации является увеличение размеров дискретизаций некоторых входов и выхода, что приводит к уменьшению размеров её интервалов, следовательно, меньшее число входов и выходов попадёт в один интервал дискретизации. А значит, меньшее число выходов будет соответствовать одному вектору входов $v(j)$ и диапазон значений дискретизированного выхода для данного входа будет меньше. Тем самым точность модели возрастёт, а размер модели увеличится. Экспериментируя с размерами дискретизации (изменяя размеры дискретизаций входов и выхода), можно подобрать более приемлемую комбинацию точности и размера модели. В частности, при наличии функциональной зависимости нередко удаётся получить дискретизацию с количеством различных (несовпадающих) выходов, не превосходящим 1 или 2.

Рассмотрим исходные данные, дискретная модель которых служит основой приближённой модели. Значение выхода такой модели определяется следующим образом. Для каждого дискретного вектора входа v_i центр тяжести выхода модели находится среди экспериментальных данных от $\min\limits_{y \in I(Y)} y + (u(U_i) - 1)\Delta(y)$ до $\min\limits_{y \in I(Y)} y + u(U_i)\Delta(y)$. В качестве значения выхода приближённой модели возьмём середину этого интервала:

$$\min_{y \in I(Y)} y + u(U_i)\Delta(y) - 0{,}5\Delta(y).$$

Результаты проведённого анализа представим совокупностью временных рядов

$$x(j),\, y(j),\, v(j),\, \tilde{y}(j),\, \tilde{w}(j), \quad j = 1, \ldots, N,$$

где $x(j)$, $y(j)$ — значения входа и выхода исходных данных; $v(j)$ — значение дискретного входа;

$$\tilde{y}(j) = \min_{y \in I(Y)} y + u(U_i)\Delta(y) - 0{,}5\Delta(y),$$

— значение выхода приближённой модели;

$$\tilde{w}(j) = |\tilde{y}(j) - y(j)| \Big[\Big(\max_{y \in I(Y)} y - \min_{y \in I(Y)} y \Big) \Big/ 100 \Big]$$

— отклонение значения выхода этой модели от значения выхода эксперимента (в процентах).

Анализ данных в случае серии экспериментов. В общей постановке задача может рассматриваться как объединение в одну модель M двух дискретных моделей: ранее построенной M_0 и построенной в результате дополнительных испытаний M_1.

Обработка данных производится следующим образом. Предполагается, что границы областей значений входов и выхода, а также размеры дискретизаций сохраняются во всех экспериментах серии. Для каждого очередного эксперимента строится дискретная модель M_1 (как в случае одного эксперимента). После этого модели объединяются в одну общую модель.

Пусть $V_0 = \{v_1, \ldots, v_p\}$ — множество векторов входов ранее построенной модели M_0. Напомним, что каждый вектор входа встречается в модели только 1 раз. Каждому v_i, $i = 1, \ldots, p$, из V_0 соответствует подпоследовательность U_i^0 значений выхода.

Пусть $V_1 = \{v_1^1, \ldots, v_q^1\}$ — множество векторов входов дискретной модели M_1. Множество входов V_1 представим в виде объединения $V_1 = V_{11} \cup V_{12}$, где $v_i \in V_{11}$, если $v_i \in V_0 \cap V_1$, и $v_i \in V_{12}$, если $v_i \notin V_0$. Равенство $V_0 \cap V_{11} = V_{11}$ справедливо только для входов V_{11} и в общем случае не выполняется для выходов U_{11}. Обозначим через U_{11}^0 множество выходов модели M_0 с входами, совпадающими с входами из V_{11}. Тогда для $v_i \in V_{11}$ множеству выходов присваивается значение $U_i^1 := U_{11} \cup U_{11}^0$. Определим координату центра тяжести дискретизированного выхода для входа v_i:

$$u(U_i^1) = \Big\{ \sum_{u_k \in U_{11}} u_k s_k + \sum_{u_k \in U_{11}^0} u_k s_k \Big\} \Big/ \Big\{ \sum_{u_k \in U_{11}} s_k + \sum_{u_k \in U_{11}^0} s_k \Big\}$$

и величину отклонения значения выхода приближённой модели от значения выхода эксперимента.

Дискретная модель с входами V_{12} строится с использованием метода для случая одного эксперимента.

Объединённая модель будет включать следующие составляющие: часть модели M_0 (исключая подмножество входов, совпадающих с входами из V_{11}) и две части модели M_1 с входами из V_{11} и V_{12}. В итоге такая модель вместе с составляющими её частями и набором характеризующих их параметров переименовывается в M_0.

Заключение. В предлагаемой работе рассмотрен способ определения существования и моделирования функциональной зависимости данных наблюдений, основанный на их дискретизации и графическом анализе. Рассмотрен также случай серии экспериментов. Предложен способ оценки модели путём вычисления отклонений значений выходов приближённой модели от значений выходов эксперимента. Разработаны (на языке Си) экспериментальные программы реализации предложенного способа обработки данных.

СПИСОК ЛИТЕРАТУРЫ

1. **Пивкин В. Я.** Построение нечетких моделей динамических объектов по данным наблюдений // Автометрия. 1998. № 3. С. 62–67.
2. **Pivkin V. Ya.** Synthesizing and correcting fuzzy model rules via series of experiments // Proc. of the IASTED Intern. Conf. Automation, Control, and Information Technology (ACIT 2002). Anaheim — Calgary — Zurich: ACTA Press, 2002. P. 164–166.

Поступило в редакцию 21 декабря 2009 г.