ОБНАРУЖЕНИЕ ЭМПИРИЧЕСКИХ ЗАКОНОМЕРНОСТЕЙ С ПОМОЩЬЮ ЭВМ
(Вычислительные системы)
1984 год Выпуск 102

УДК 517.8:33

ОБ ИССЛЕДОВАНИИ ПОВЕДЕНИЯ ЗАКОНОМЕРНОСТЕЙ ПРИ ИСПОЛЬЗОВАНИИ РАЗЛИЧНЫХ МЕТРИК

А.Н.Антамошкин, О.М.Тимофеева

В работе [1] для управления донозологическим контролем здоровья популяции в условиях крупного химического производства предложен алгоритм, использующий выявленные закономерности. Вкратце, алгоритм состоит в следующем.

Пусть $X = (x_1,\dots,x_n)$ - вектор управляемых параметров, $Y = (y_1,\dots,y_m)$ - вектор контролируемых, но неуправляемых параметров. Имеется статистика из T пар $\{X^t,Y^t\}$ и соответствующих значений критерия качества управления $Z(X^t, Y^t)$, $t = \overline{1,T}$. Требуется для заданных Y^{T+k} найти управления $X^{T+k} \in \Omega$, где Ω - множество допустимых управлений, таких, чтобы значения критерия качества $Z(X^{T+k},Y^{T+k})$ были минимальными $(k = 1,2,\dots)$.

По имеющейся статистике строится опорное множество $L = \{(X^i,Y^i): Z(X^i,Y^i) \le C,\ i = \overline{1,N}, N \le T\}$, где $C = \mathrm{const}$ заранее задаваемое значение критерия (при $Z \le C$ управление считается удовлетворительным). Оптимальное управление X^{T+k} для заданного Y^{T+k} $(k = 1,2,\dots)$ находится из условия

$$\sum_{i=1}^{N} |\rho(Y^i,Y^{T+k}) - W_L\rho(X^i,X)| \to \min_{X\in\Omega},$$

где ρ - заданная функция расстояния (метрика) [2], а $W_L = (\sum_{p,q=1} |\rho(Y^p,Y^q)-\rho(X^p,X^q)|)/C_N^2$ - коэффициент пропорциональности (C_N^2 - число сочетаний).

В решавшихся с помощью этого алгоритма задачах предупреждения профессиональных заболеваний [1] среди компонент векторов X

и Y встречалось три типа переменных - булевы, целочисленные, непрерывные. С помощью методики, предложенной в [3], осуществлялось погружение векторов X и Y в R^n, и в алгоритме использовалась обычная евклидова метрика - l_2 [2]. Однако выбор в качестве функции расстояния именно этой метрики был основан только на том,что евклидова метрика очень популярна и наиболее употребительна. В [2] приведены примеры других достаточно часто применяемых для R^n функций расстояния. По-видимому, качество управления должно зависеть от того, какая мера используется в алгоритме в качестве функции расстояния. Для выяснения этого вопроса были проведены экспериментальные исследования.

Эксперименты проводились для случая, как предполагалось,наиболее благоприятного для евклидовой метрики - компоненты векторов X и Y изменялись непрерывно. Рассмотрим один из экспериментов: $X^i \in (0,\infty)$, $Y^i \in R^{13}$, $i = \overline{1,11}$, причем компоненты Y по i изменялись очень неравномерно, так, если $y_6 \in [0,I;\ 0,4]$, то $y_{13} \in [200,600]$, $i = \overline{1,11}$, $T = 10$, $N = 4$. Физически значениям компонент вектора Y^i соответствовали значения биологических маркеров i-го индивидуума. В качестве X выступал хронотип, измеряемый в баллах. Критерию управления соответствовала гомеостатическая устойчивость индивидуума. Оптимальное управление X^{11} определялось шесть раз. При этом в качестве функции расстояния использовались [2,4,5] ($X,Y \in R^n$):

$\rho_1(X,Y) = \sum_{j=1}^{n} |x_j - y_j|$ - l_1- норма ($X^{11} \approx 6,7$);

$\rho_2(X,Y) = (\sum_{j=1}^{n} |x_j-y_j|^2)^{1/2}$ - l_2- норма (евклидова метрика) ($X^{11} \approx 6,75$);

$\rho_3(X,Y) = (\sum_{j=1}^{n} |x_j-y_j|^3)^{1/3}$ - l_3- норма ($X^{11} \approx 6,8$);

$\rho_4(X,Y) = (\sum_{j=1}^{n} |x_j-y_j|^4)^{1/4}$ - l_4- норма ($X^{11} \approx 6,85$);

$$\rho_5(X,Y)=\left(\frac{1}{n}\sum_{j=1}^{n}\left(\frac{x_j-y_j}{x_j-y_j}\right)^2\right)^{1/2}$$ – коэффициент дивергенции $(X^{11}\approx 6{,}5)$;

$$\rho_6(X,Y)=\left(\sum_{j=1}^{n}(\sqrt{x_j}-\sqrt{y_j})^2\right)^{1/2}$$ – мера Джеффриса-Матуситы $(X^{11}\approx 5{,}5)$.

Эвристические меры отдаленности ρ_5 и ρ_6 не являются метриками по классическому определению, но часто применяются на практике, в том числе и для $X,Y \in R^n$ [2].

В действительности оптимальное управление $X^{11}=7$, интервал допустимых управлений X^{11} – [6,7;7,I].

Таким образом, анализ полученных результатов показывает, что все управления, найденные в качестве оптимальных при использовании метрик l_p-норма (p=1,4), принадлежат интервалу допустимых управлений; точность решения растет с ростом p (наиболее близкое к истинному значению оптимального. управление значение получено при p=4); использование мер ρ_5 и ρ_6 дает неудовлетворительные результаты.

Аналогичные результаты наблюдались и в других экспериментах (для них, к сожалению, были известны только интервалы допустимых управлений, истинные значения оптимальных управлений физически определить не удалось) – использование метрик l_p-нормы $(p=\overline{1,4})$ давало значения оптимальных управлений, принадлежащие интервалу допустимых управлений, причем с ростом p значения найденных управлений смещались к середине интервала, использование в алгоритме в качестве функций расстояния коэффициента дивергенции и меры Джеффриса-Матуситы давало, как правило, неудовлетворительные результаты.

Предварительно можно сделать следующие выводы.

При управлении, основанном на выявлении закономерностей, оправдано использование в качестве функции расстояния метрик l_p-нормы, причем с ростом p повышается качество управления. По-видимому, имеет смысл продолжить исследования для случая $p > 4$.

Употребление в качестве функций расстояния коэффициента дивергенции и меры Джеффриса-Матуситы (по крайней мере, в случае переменных, замеряемых в непрерывных шкалах) нецелесообразно,что, возможно, объясняется тем, что первоначально они были предложены для специальных задач [4,5].

Представляют интерес исследования качества управления при использовании рассмотренных функций расстояний в случае разношкальных переменных, а также при использовании других метрик, например, сюпремум-нормы, Махаланобиса [2].

Литература

I. АНТАМОШКИН А.Н., КУРИС Е.В., ЧЕКМЕНЕВ В.А. Управление донозологическим контролем здоровья популяции. -В кн.: Динамика систем (Оптимизация и адаптация). Горький, I98I, с. 226-230.

2. ДЮРАН Б., ОДЕЛЛ П. Кластерный анализ. -М.: Статистика, I977.

3. СТОЯН Ю.Г. Об одном отображении комбинаторных множеств в евклидово пространство. -Харьков, I982. (Препринт/ Институт проблем машиностроения АН УССР: I73).

4. CLARK P.I. An extension of the coefficient of devergence for use with multiple characters.-Copeia,1952,N 2,p.61-64.

5. OICE L.R. Measures of the amount of ecological associa - tion between species.-Ecology,1945,N 26,p.297-302.

Поступила в ред.-изд.отд.
22 марта I984 года