 Közlemények 31/1984 153-158

МОНОТОННЫЕ СИСТЕМЫ НА МАТРИЦАХ ДАННЫХ

Е.Н. Кузнецов, И.Б. Мучник, F. Хенчей, Н.Ф. Чкуасели

1. Монотонные системы /МС/ для агрегирования данных /в частности, для решения задач типа задачи классификации множества объектов/ предложены И.Э. Муллатом в 1971 г. [1, 2]. Метод МС характеризуется большой общностью подхода к задачам такого типа, но в отличии от других методов требует задания числовой функции связи между отдельным элементом и любым подмножеством исходного множества - функции связи "элемент-подмножество" /ФСЭП/. Сам метод МС средств порождения таких функций, а тем самым, различных монотонных систем не содержит. В практических исследованиях [3, 4] использовалось лишь несколько конкретных МС. В данной работе ставится задача показать регулярные способы построения монотонных систем для реализации различных макроописаний матрицы данных и создания единого программного обеспечения для такого анализа.

2. МС и задача структуризации. Монотонной системой $\langle W,\Pi \rangle$ называется множество W, $|W|=N$, с заданной на нем функцией /ФСЭП/ $\Pi(i,H)$, $i\in H$, $H\subseteq W$, удовлетворяющей неравенству $\Pi(i,H\setminus j)\leq\Pi(i,H)$, $\forall i\in H\setminus j$, $\forall j\in H$, $\forall H\subseteq W$ /свойство монотонности/ [2]. Ядром МС называется множество $G\leq W$, для которого $F(G)=\max_{H\leq W} F(H)$, где $F(H)=\min_{i\in H} \Pi(i,H)$ [1].

Для выделения наибольшего по мощности ядра МС предложены эффективные алгоритмы [1, 2, 5]. В ходе работы некоторых из них строится последовательность вложенных множеств $\bar{\Gamma} = \langle\Gamma_1,\Gamma_2,\ldots,\Gamma_p\rangle$, $\Gamma_1 = W$, $\Gamma_p = G$, являющихся локальными максимумами функции $F(H)$ [2].

Для структуризации исходного множества W используется как разбиение его на две части: G и $W\setminus G$, так и на подмножества $\Gamma_1\setminus\Gamma_2$, $\Gamma_2\setminus\Gamma_3,\ldots,\Gamma_{p-1}\setminus\Gamma_p$, Γ_p . Кроме того, если функция $\Pi(i,H)$

имеет смысл расстояния, то G можно рассматривать как множество эталонов-классообразующих элементов искомой классификации. Разбиение W на "однородные" части получают также путем последовательного построения МС$\langle W \setminus G, \Pi \rangle$ на множестве $W \setminus G$ элементов, не вошедших в ядро предыдущей системы, затем на оставшихся вне нового ядра и т.д. /до исчерпания исходного множества/.

3. Выбор множества W элементов МС на матрице данных $\Phi = \| \varphi_{ij} \|$ (NxM). Есть четыре возможности [6]: а/ множество X объектов - строк Ф; б/ множество Y признаков - столбцов Ф; в/ $X \cup Y$; г/ множество пар (i,j), $i = \overline{1,N}$, $j = \overline{1,M}$. При этом произвольному подмножеству H множества W /в т.ч. ядру G/ соответствует часть Ф(H) матрицы Ф определенной конфигурации: а/ горизонтальная полоса, б/ вертикальная полоса, в/ "крест" или его прямоугольный блок - пересечение, г/ "пятно" произвольной формы либо минимальный охватывающий его прямоугольный блок.

Эти четыре способа можно использовать и комбинированно, например, сначала представив матрицу разделенной на "однородные" вертикальные полосы, а затем - внутри каждой из них независимо - на горизонтальные блоки, или наоборот. /Можно менять характер элементов МС и после каждого выделения ядра, получая макроописание матрицы типа "паркет./

Вопрос о связи МС на множестве X, Y и $X \cup Y$ частично решают следующие теоремы. Введем обозначения: $\forall H \subseteq X \cup Y$, $H_X = H \cap X$, $H_Y = H \cap Y$; $\Phi(H_Y)$ - подматрица матрицы Ф, определяемая столбцами H_Y; $\Pi(i,H,\Phi)$ - ФСЭП $\Pi(i,H)$, вычисляемое по матрице Ф.

Теорема 1. Если G - ядро МС $\langle W, \Pi^1 \rangle$, где

$$\Pi^1(i,H,\Phi) = \begin{cases} \Pi(i,H_X,\Phi), & \text{если} \quad i \in H_X \subseteq X, \\ \Pi(i,H_Y,\Phi^T), & \text{если} \quad i \in H_Y \subseteq Y, \end{cases}$$

а G^X и G^Y - ядра, соответственно, МС $\langle X,\Pi,\Phi\rangle$ и $\langle Y,\Pi,\Phi^T\rangle$, то верно одно из трех

а/ $G=G^X$, $G^Y\subseteq W\backslash G$, если $F(G^X)>F(G^Y)$,

б/ $G=G^Y$, $G^X\subseteq W\backslash G$, если $F(G^X)<F(G^Y)$,

в/ $G=G^X\cup G^Y$, если $F(G^X)=F(G^Y)$.

<u>Теорема 2.</u> Если для ядра G^X МС $\langle X,\Pi,\Phi\rangle$ на множестве строк X матрицы Φ выполняется $F(G^X,\Phi)\leq F(Y,\Phi^T(G^X))$, где

$$F(G^X,\Phi)=\min_{i\in G^X}\Pi(i,G^X,\Phi),\quad F(Y,\Phi^T(G^X))=\min_{j\in Y}\Pi(j,Y,\Phi^T(G^X)),$$ то

$G_X=G\cap X=G^X$, где G - ядро МС $\langle W,\Pi^1\rangle$, удовлетворяющей

$$\Pi^1(i,H,\Phi)=\begin{cases}\Pi(i,H_X,\Phi(H_Y)), & \text{если } i\in H_X\subseteq X,\\ \Pi(i,H_Y,\Phi^T(H_X\backslash j)), & \text{если } i\in H_Y\subseteq Y.\end{cases}$$

Таким образом, МС на $W=X\cup Y$ при нектороных условиях можно конструировать из "частных" МС, определенных на X и Y отдельно, либо как МС только на множестве X' строк или Y' столбцов другой матрицы, составленной из исходной матрицы Φ и матрицы Φ^T, полученной транспорированием исходной. В случае г/ исходную матрицу Φ можно интерпретировать как двудольный граф с весами и строить МС на дугах графа 1, 7.

4. <u>Три типа функций связи "элемент-подмножество".</u>

1/ суммарно-попарные ФСЭП $\Pi(i,H)=\sum_{k\in H}a_{ik}$, где a_{ik} - мера парной связи i-го и k-го элементов, в частности один из известных коэффициентов связи двух векторов /коэффициент корреляции признаков, число общих признаков объектов, расстояние Хэмминга и др./ [1, 5]; 2/ представительские ФСЭП $\Pi(i,H)=a_{ih}$, где a_{ih} - мера парной связи, а h - элемент, быть может, специально сконструированный, - "представитель" множества H /например, для булевой матрицы Φ признаки элемента - представителя h могут определяться как $\varphi_{hj}=\bigcap_{k\in H}\varphi_{kj}$ или $\varphi_{hj}=\bigcup_{k\in H}\varphi_{kj}$ [3];

3/ суммарно-признаковые ФСЭП $\Pi(i,H) = \sum_{j\in y_i} \omega_j(H)$, где $\omega_j(H)$ - вес признака j на множестве элементов H, y_i - множество признако i-го объекта /например, $\omega_j(H) = \sum_{k\in H} \varphi_{kj}$ - число объектов из H, имеющих этот признак [6]. На компоненты ФСЭП каждого типа накладываются ограничения, необходимые для выполнения свойства монотонности /например, в 1/ - $a_{ik} \geq 0$, в 3/ - $\omega_j(H\backslash k) \leq \omega_j(H)$. В работе рассматривается целая "коллекция" конкретных примеров ФСЭП каждого типа.

Границы между ФСЭП разных типов всегда жесткие.

Пример. Пусть $a_{ik} = \sum_{j\in Y} \varphi_{ij} \cdot \varphi_{kj}$ - число общих признаков пары объектов /Φ - булева/, а $\omega_j(H) = \sum_{k\in H} \varphi_{kj}$ - число объектов в H с j-ым признаком. Тогда $\Pi^{1)}(i,H) = \sum_{k\in H} a_{ik}$ и $\Pi^{2)}(i,H) = \sum_{j\in y_i} \omega_j(H)$ совпадают.

5. Порождение параметрических семейств монотонных систем.

Теорема 3. Если $\langle W,\Pi_1\rangle$ и $\langle W,\Pi_2\rangle$ - МС, то $\langle W,\Pi\rangle$ - тоже МС, если

$$\Pi(i,H) = \alpha_1 \Pi_1(i,H) + \alpha_2 \Pi_2(i,H), \quad \alpha_1,\ \alpha_2 \geq 0.$$

Теорема 4. Если $\langle W_1 \Pi_1\rangle$ и $\langle W_1 \Pi_2\rangle$ - МС, то $\langle W,\Pi\rangle$ - тоже МС, если

$$\Pi(i,H) = [\Pi(i,H)]^{\beta_1} \cdot [\Pi_2(i,H)]^{\beta_2}, \quad \beta_1,\ \beta_2 \geq 0.$$

Очевидны обобщения на произвольное число базовых МС. Возможна и внутренняя параметризация, например,

$$\Pi(i,H) = \sum_{k\in H} a_{ik}^{\gamma} \quad \text{или} \quad \Pi(i,H) = \sum_{j\in Y} [\omega_j(H)]^{\delta}, \quad \gamma,\delta \geq 0$$

Легко видеть, таким образом, что в ФСЭП I-го и III-го типов можно использовать не сумму, а любую симметричную по аргументам, т.е. a_{ik} или $\omega_j(H)$, функцию, например,

$$\Pi(i,H) = \sum_{s\in H} \alpha_s \cdot \prod_{k\in H} \alpha_k \cdot (a_{ik})^{\gamma_s}$$

Таким образом, имея разные базовые ФСЭП трех типов и варьируя значения параметров, можно строить разнообразные МС с наперед заданными содержательными свойствами и, тем самым, получать разные макроописания исходных данных.

ЛИТЕРАТУРА

1. Муллат И.Э.: Экстремальные подсистемы монотонных систем I-III.- Автоматика и телемеханика, 1976, №5, с. 130-139; №8, с. 169-178; 1977, №1, с. 109-119.

2. Кузнецов Е.Н., Мучник И.Б., Шварцев Л.В.: Монотонные системы и их свойства. - В кн.: Анализ нечисловой информации в социологических исследованиях. М.: Наука, 1984, с. 123-149.

3. Кузнецов Е.Н., Мучник И.Б.: Анализ распределения функций в организационной системе. - Автоматика и телемеханика, 1982, №10, с. 119-127.

4. Выханду Л.К., Выханду П.Л.: Быстрый поиск на битматрицах. - В кн.: Труды Таллинского политехн. ин-та, 1983, №554, с. 49-60.

5. Кузнецов Е.Н.: Анализ структуры матрицы связей с помощью построения на ней монотонной системы. - Автоматика и телемеханика, 1980, №7, с. 128-136.

6. Выханду Л.К.: Монотонные системы в анализе данных. - В кн.: Труды Таллинского политехн. ин-та, 1981, №511, с. 91-100.

7. Мучник И.Б.: Обработка экспертных суждений о структурных объектах. - В кн.: Экспертные оценки в задачах управления. Ин-т проблем управления, 1982, с. 27-32.

S U M M A R Y

E.N. Kuznetzov, I.B. Muchnik, G. Hencsey, N.F. Tchkuasely

MONOTONIC SYSTEMS ON DATA MATRICES

In the paper a technique for the structuralization of large-scale data matrices is considered, namely the use of the theory of monotonic systems. This is expected to be highly suitable for different applications.

For the method of monotonic systems a numerical function coupling any element with any subject of the basic set is needed.

We discuss three main classes of such functions and four possibilities of determining them for data matrices.

An idea of a generator for parameter families of such functions defining monotonic systems is suggested.

Ö S S Z E F O G L A L Á S

E.N. Kuznetzov, I.B. Muchnik, G. Hencsey, N.F. Chkuasely

MONOTON RENDSZEREK ADAT-MÁTRIXOKON

A cikkben a szerzők a nagyméretü adat-mátrixok strukturalizálására adnak meg egy technikát, amely a monoton rendszerek módszerén alapszik. Ez várhatóan igen hasznos lesz különböző alkalmazásokra.

A monoton rendszerek módszere egy pont-halmaz függvény megadásán alapszik.

A szerzők ilyen függvényeknek három osztályát, illetve adat-mátrixokon való négy fajta megadását tárgyalják. Ezen függvények paraméteres családjainak generálását is megemlitik.