УДК 519.68

С.М. АГАЯН, Ш.Р. БОГОУТДИНОВ, М.Н. ДОБРОВОЛЬСКИЙ

ДИСКРЕТНЫЕ СОВЕРШЕННЫЕ МНОЖЕСТВА И ИХ ПРИМЕНЕНИЕ В КЛАСТЕРНОМ АНАЛИЗЕ

Аннотация. Описаны исследования в рамках дискретного математического анализа формализации нечеткого понятия «кластер». Предпринята попытка математически реализовать на базе так называемых дискретных совершенных множеств эвристическое определение Эверитта. Рассмотрены дискретные совершенные множества (варианты кластера) и на их основе построен алгоритм DPS, осуществляющий фильтрацию исходного пространства путем выделения в нем максимального плотного на общем фоне подмножества.

Ключевые слова: дискретный математический анализ, плотность, радиус локализации, нечеткое множество, совершенное множество, морфологический фильтр.

ВВЕДЕНИЕ

Настоящая статья является продолжением исследований в рамках дискретного математического анализа (ДМА) формализации нечеткого понятия «кластер» с последующим построением основанных на ней алгоритмов кластерного анализа, например, Роден [1–3], Кристалл [4, 5] и Монолит [5, 6]. В работе предпринята попытка математически реализовать эвристическое определение Эверитта [7], согласно которому «кластеры — это «непрерывные» области (некоторого) пространства с относительно более высокой плотностью точек, отделенные от других таких же областей областями с относительно низкой плотностью точек». Один из возможных подходов к решению этой задачи состоит в следующем: сначала формализуется понятие плотного на общем фоне множества, далее в исходном пространстве выделяется максимальное такое множество, которое, в свою очередь, разбивается на компоненты связности. Последние будут плотными, изолированными одна от другой областями в исходном пространстве, т.е. кластерами. Все упомянутые выше алгоритмы, каждый по-своему, осуществляют эту программу до этапа разбиения на связные компоненты, выделяя максимальные интересные (реперные) плотные подмножества в исходном пространстве и тем самым очищая его от фона или производя топологическую фильтрацию.

Последний этап разбиения на кластеры можно осуществить как традиционным кластерным анализом [7–9], так и ДМА. Полные реализации определения Эверитта представляют собой новый, посткластеризационный этап в кластерном анализе, поскольку они не только разбивают исходное пространство на однородные части, но и предварительно очищают (фильтруют) его. Примерами таких алгоритмов являются дополненные кластерным разбиением алгоритмы Роден, Кристалл и Монолит.

Цель настоящей статьи — определение и изучение дискретных совершенных множеств (вариантов кластера), а также построение на их основе алгоритма DPS (Discrete Perfect Sets), осуществляющего фильтрацию исходного пространства путем выделения в нем максимального плотного на общем фоне подмножества. В этом смысле алгоритм DPS продолжает ДМА-алгоритмы Роден, Кристалл и Монолит.

КОНСТРУКЦИЯ $A(\alpha)$

Пусть X — конечное множество, $A, B, \ldots$ — подмножества и $x, y, \ldots$ — точки в нем.

Определение 1. Назовем плотностью P на множестве X отображение из $2^X \times X$ в отрезок $[0,1]$, возрастающее по первому аргументу

$$P(A, x) = P_A(x) \ \ \forall x \in X;$$
$$A \subset B \Rightarrow P_A(x) \le P_B(x), \tag{1}$$

где $P_A(x)$ — плотность подмножества A в точке x.

При фиксированном x функция $P_A(x)$ есть нечеткая мера на X [10], так что плотность P — семейство нечетких мер на X, параметризованное множеством X.

Для $\alpha \in [0, 1]$ и подмножества $A \subseteq X$ с помощью плотности P определим последовательность оболочек плотности (α-n-оболочек плотности P) A в X:

$$A^1(\alpha) = A^1 = \{x \in X : P_A(x) \ge \alpha\},$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$A^n(\alpha) = A^n = \{x \in X : P_{A \bigcup A^{n-1}}(x) \ge \alpha\}.$$

Комментарий 1. Плотность $P_A(x)$ трактуется как мера предельности A в точке x. Если она значительна ($P_A(x) \ge \alpha$), то точка x считается предельной для A. Таким образом, первая оболочка A^1 является множеством всех в этом понимании предельных для A точек в X. Точки из второй оболочки A^2 будут в общем случае предельными для A в X через A^1, т.е. второго рода, и т.д.

Утверждение 1. Имеют место включения

$$A^1 \subseteq A^2 \subseteq \ldots \subseteq A^n \subseteq \ldots \tag{2}$$

Доказательство по индукции. Начало $n = 1$: если $x \in A^1$, то $P_A(x) \ge \alpha \Rightarrow$ $\Rightarrow P_{A \bigcup A^1}(x) \ge P_A(x) \ge \alpha \Rightarrow x \in A^2$ (следствие монотонности (1)).

Индукционный шаг $n-2 \to n-1$: предположение $A^{(n-2)} \subseteq A^{(n-1)}$.

Доказательство перехода $n-1 \to n$: если $x \in A^{(n-1)}$, то $P_{A \cup A^{(n-2)}}(x) \ge \alpha$. По предположению $A^{(n-2)} \subseteq A^{(n-1)}$, следовательно, $A \bigcup A^{(n-2)} \subseteq A \bigcup A^{(n-1)}$ и поэтому $P_{A \bigcup A^{(n-1)}}(x) \ge P_{A \bigcup A^{(n-2)}}(x) \ge \alpha \Rightarrow x \in A^n$. Последнее равносильно включению $A^{(n-1)} \subseteq A^n$.

В силу конечности X в неубывающей последовательности (2), начиная с некоторого номера n^*, наступит стабилизация: $A^1 \subset A^2 \subset \ldots \subset A^{n^*} = A^{n^*+1} = \ldots$ Назовем ее ∞-оболочкой (α-∞-оболочкой) A и обозначим $A^\infty = A^\infty(\alpha)$.

Множество A^∞ полуинвариантно: ее первая оболочка плотности $(A^\infty)^1$ не выходит за пределы A^∞.

Утверждение 2. Если $P_{A^\infty}(x) \ge \alpha \Rightarrow x \in A^\infty \Leftrightarrow (A^\infty)^1 \subseteq A^\infty$.

Доказательство от противного. Воспользуемся финитным представлением A^∞: $A^\infty = A^{n^*}$. Если $P_{A^{n^*}}(x) \ge \alpha$ и $x \notin A^{n^*}$, то $P_{A \bigcup A^{n^*}}(x) \ge P_{A^{n^*}}(x) \ge \alpha \Rightarrow$ $\Rightarrow x \in A^{(n^*+1)}$ и поэтому $A^{n^*} \subset A^{(n^*+1)}$. Однако по условию $A^{n^*} = A^{(n^*+1)}$. Полученное противоречие доказывает утверждение.

Отсюда непосредственно вытекает, что для множества A^∞ ряд (2) постоянен.

Следствие. Справедливо следующее соотношение**:**

$$(A^\infty)^n = (A^\infty)^1 \ \forall n \ge 2. \tag{3}$$

Доказательство. Ранее было установлено, что $(A^\infty)^1 \subseteq A^\infty$, следовательно, $(A^\infty)^2 = \{x \in X : P_{A^\infty \bigcup (A^\infty)^1}(x) = P_{A^\infty}(x) \ge \alpha\} = (A^\infty)^1$ и т.д.

Двойную бесконечную оболочку для A обозначим $A^{2\infty}$. Таким образом,

$$A^{2\infty} = (A^\infty)^\infty = (A^\infty)^1 \subseteq A^\infty. \tag{4}$$

Продолжим эти рассуждения в виде следующей схемы:

$$\begin{array}{ccccccccc}
A & \leftrightarrow & A^1 & \subseteq & A^2 & \subseteq \ldots & \leftrightarrow & A^{\infty} \\
 & & & & & & & \cup\vert \\
A^{\infty} & \leftrightarrow & (A^{\infty})^1 & = & (A^{\infty})^2 & = \ldots & = & A^{2\infty} \\
\cup\vert & & & & & & & \cup\vert \\
A^{2\infty} & \leftrightarrow & (A^{2\infty})^1 & = & (A^{2\infty})^2 & = \cdots & = & A^{3\infty} \\
\cup\vert & & & & & & & \cup\vert \\
\vdots & & & & & & & \vdots \\
\cup\vert & & & & & & & \cup\vert \\
A^{m\infty} & \leftrightarrow & (A^{m\infty})^1 & = & (A^{m\infty})^2 & = \ldots & = & A^{(m+1)\infty} \\
\cup\vert & & & & & & & \cup\vert \\
\vdots & & & & & & & \vdots
\end{array} \qquad (5)$$

В силу конечности X в невозрастающей последовательности $A^{\infty} \supseteq A^{2\infty} \supseteq \ldots$ $\ldots \supseteq A^{m\infty} \supseteq \ldots$, начиная с некоторого номера m^*, наступит стабилизация $A^{\infty} \supset \ldots$ $\ldots \supseteq A^{m^*\infty} = A^{(m^*+1)\infty} = \ldots$

Определение 2. Множество $A^{m^*\infty}$ обозначим $A(\alpha)$ и назовем α-оболочкой A в X на основе плотности P:

$$A(\alpha) = \bigcap_m A^{m\infty}(\alpha). \qquad (6)$$

Итак, процесс построения $A(\alpha)$ имеет стадию возрастания от $A^1(\alpha)$ до $A^{\infty}(\alpha)$ и стадию убывания от $A^{\infty}(\alpha)$ до $A(\alpha)$:

$$A \to A^1(\alpha) \subseteq A^2(\alpha) \subseteq \ldots A^{\infty}(\alpha) \supseteq A^{2\infty}(\alpha) \supseteq \ldots \supseteq A(\alpha). \qquad (7)$$

Утверждение 3. Множество $A(\alpha)$ совпадает со своей первой оболочкой $(A(\alpha))^1 = A(\alpha)$.

Доказательство. Согласно (4) множество $A(\alpha) = A^{m^*}(\alpha)$, а также $(A(\alpha))^1 = A^{(m^*+1)}(\alpha)$.

Следствие. Множество $A(\alpha)$ совпадает со своей α-оболочкой:

$$(A(\alpha))(\alpha) = A(\alpha). \qquad (8)$$

Последнее означает, что оператор построения $A(\alpha)$ по A в пространстве X является идемпотентным.

Комментарий 2. Утверждение 2 означает, что A^{∞} является замкнутым множеством, потому что содержит все свои α-предельные в X точки. Движение внутри A^{∞} приводит к α-совершенному подмножеству $A(\alpha)$, поскольку утверждение 3 означает, что множество $A(\alpha)$ состоит ровно из тех точек в X, которые к нему α-предельны, т.е. ровно из тех точек, где плотность $A(\alpha)$ больше или равна α :

$$A(\alpha) = \{x \in X : P_{A(\alpha)}(x) \geq \alpha\}. \qquad (9)$$

Во всех точках дополнения плотность $A(\alpha)$ меньше α:

$$\overline{A(\alpha)} = \{x \in X : P_{A(\alpha)}(x) < \alpha\}. \qquad (10)$$

Многочисленные исследования и приведенные далее примеры показывают, что условия (9) и (10) можно считать одним из вариантов формального понимания свойства «быть плотным на общем фоне», множества $A(\alpha)$ — плотными в пространстве X, а процесс их получения $A \to A(\alpha)$ (7) — еще одной ДМА-реализацией определения Эверитта.

ОПРЕДЕЛЕНИЕ АЛГОРИТМА DPS

Процесс построения для множества A в универсуме X на основе плотности P α-оболочки $A(\alpha) = A_P(\alpha | X)$ (7) называется алгоритмом DPS:

$$\text{DPS}(\cdot) = \text{DPS}(\cdot | X, P, \alpha): 2^X \to 2^X .$$

Согласно (8) алгоритм DPS идемпотентен ($\text{DPS}^2 = \text{DPS}$). Неподвижные относительно него подмножества называются α-совершенными множествами (α-DPS-множествами) в X:

$$A \text{ — } \alpha\text{-DPS-множество в } X \Leftrightarrow \text{DPS}(A | X, P, \alpha) = \\ = A \Leftrightarrow A = \{x \in X : P_A(x) \geq \alpha\} = A^1. \qquad (11)$$

В общем случае, как было установлено ранее (7), алгоритм DPS имеет две стадии: возрастающую $A^n \uparrow A^\infty \Leftrightarrow A \to A^1 \subseteq \ldots = A^\infty$ и убывающую $A^{m\infty} \downarrow A(\alpha) \Leftrightarrow$ $\Leftrightarrow A^\infty \supseteq A^{2\infty} \supseteq \ldots = A(\alpha)$.

Существуют ситуации, в которых алгоритм DPS «работает быстрее» и имеет не более одной стадии. Тривиальный случай приведен в (11): DPS неподвижен на $A \equiv$ ноль стадий алгоритма DPS на $A \equiv A$ совершенно. Исследуем DPS с одной стадией.

Возрастающий DPS. Имеется только возрастающая стадия $A^n \uparrow A^\infty$, убывающей стадии не существует, следовательно, $(A^\infty)^1 = A^\infty$, т.е. A^∞ α-совершенно. Итак, DPS имеет только возрастающую стадию на $A \Leftrightarrow A^\infty(\alpha)$ и является α-совершенным.

Пример 1. Пусть $A \subset A^1$, тогда

$$A^2 = \{x \in X : P_{A \bigcup A^1}(x) \geq \alpha\} = \{x \in X : P_{A^1}(x) \geq \alpha\} = (A^1)^1.$$

Далее аналогично

$$A^{i+1} = \{x \in X : P_{A \bigcap A^i}(x) \geq \alpha\} = (A^i)^1 = (A^1)^i.$$

Если $A^\infty = A^{n*}$, то $A^{n*} = A^{(n*+1)} = (A^{n*})^1$, т.е. A^∞ α-совершенно.

Далее приведены также и другие примеры возрастающего DPS.

Убывающий DPS. Имеется только убывающая стадия $A^{m\infty} \downarrow A(\alpha)$, возрастающей стадии не существует, следовательно, $A^1 = A^\infty$. Приведем простой и эффективный критерий этой ситуации.

Утверждение 4. Справедлива следующая эквивалентность: $A^1 = A^\infty \Leftrightarrow A^1 = A^2$.

Доказательство. Достаточность очевидна, поскольку $A^1 \subseteq A^2 \subseteq A^\infty$.

Необходимость. Если $A^1 = A^2$, то $A^3 = \{x \in X : P_{A \bigcup A^2}(x) \geq \alpha\} = \{x \in X : P_{A \bigcup A^1}(x) \geq \alpha\} = A^2$. Аналогично $A^4 = A^3, \ldots, A^{i+1} = A^i$ и т.д.

Таким образом, алгоритм DPS имеет только убывающую стадию на $A \Leftrightarrow A^1 = A^2$.

Пример 2. Пусть $A^1 \subset A$, тогда

$$A^2 = \{x \in X : P_{A \bigcup A^1}(x) \geq \alpha\} = \{x \in X : P_A(x) \geq \alpha\} = A^1.$$

Очевидно, что условие $X^1 \subset X$ выполнено, поэтому на всем пространстве X алгоритм DPS всегда убывающий. Ввиду большой практической значимости этого алгоритма приведем его полностью:

$$X(\alpha) = \bigcap X^i(\alpha); \;\; X^i(\alpha) = \{x \in X : P_{X^{i-1}(\alpha)}(x) \geq \alpha\}, \;\; i \geq 1; \;\; X^0(\alpha) = X.$$

Конструкция $A(\alpha)$ зависит от четырех составляющих: непосредственно пространства X, множества A, плотности P и уровня α: $A(\alpha)=A_P(\alpha \mid X)$.

Установим, что зависимости по A, P и X являются возрастающими, а зависимость по α — убывающей.

В доказательствах приведенных далее утверждений 5–8 использованы схема и обозначения (7).

Утверждение 5. Если $A \subseteq B$, то $A(\alpha) \subseteq B(\alpha)$ (рис. 1).

Доказательство. С помощью индукции покажем вначале, что $A^n \subseteq B^n$.

Начало $n=1$: если $x \in A^1$, то $P_A(x) \geq \alpha \Rightarrow P_B(x) \geq P_A(x) > \alpha \Rightarrow x \in B^1$ (следствие монотонности (1)).

Индукционный шаг $n-2 \to n-1$: предположение $A^{(n-1)} \subseteq B^{(n-1)}$.

Доказательство перехода $n-1 \to n$: $A^n \subseteq B^n$. Пусть $x \in A^n$, тогда $P_{A \bigcup A^{(n-1)}}(x) \geq \alpha$. Однако $A \bigcup A^{(n-1)} \subseteq B \bigcup B^{(n-1)}$ вследствие начальных условий и предположения индукции, следовательно, $P_{B \bigcup B^{(n-1)}}(x) \geq P_{A \bigcup A^{(n-1)}}(x) \geq$ $\geq \alpha \Rightarrow x \in B^n$.

Начиная с конечного n имеется продолжение на бесконечность $A \subseteq B \Rightarrow$ $\Rightarrow A^\infty \subseteq B^\infty$:

$$\left.\begin{array}{ccccccc} A^1 & \subseteq & A^2 & \subseteq \ldots \subseteq & A^n & \subseteq \ldots \\ \bigcap| & & \bigcap| & & \bigcap| & \\ B^1 & \subseteq & B^2 & \subseteq \ldots \subseteq & B^n & \subseteq \ldots \end{array}\right\| \Rightarrow A^\infty = \bigcup A^n \subseteq \bigcup B^n = B^\infty .$$

Справедливо продолжение на показатели $k\infty$:

$$A^\infty \subseteq B^\infty \Rightarrow A^{2\infty} = (A^\infty)^\infty \subseteq (B^\infty)^\infty = B^{2\infty}$$

и далее для всех k. Окончательно с помощью (6) имеем

$$A(\alpha) = \bigcap A^{k\infty} \subseteq \bigcap B^{k\infty} = B(\alpha).$$

Пример 3. Универсум X — равномерная сетка 200×200 с шагом 1 по обеим осям, A (см. рис. 1, *а*) и B (см. рис. 1, *б*) — некоторые множества в $X: A \subset B$, $|A| = 7957$, $|B| = 10131$, плотность $P_A(x) = |D_A(x, r)|$ с радиусом локализации $r = 3$ и $\alpha = 0.4$. В результате работы алгоритма DPS получены два множества: $A(\alpha) \subset B(\alpha)$ ($|A(\alpha)| = 13642$, $|B(\alpha)| = 14747$, см. рис. 1, *в*, *г*).

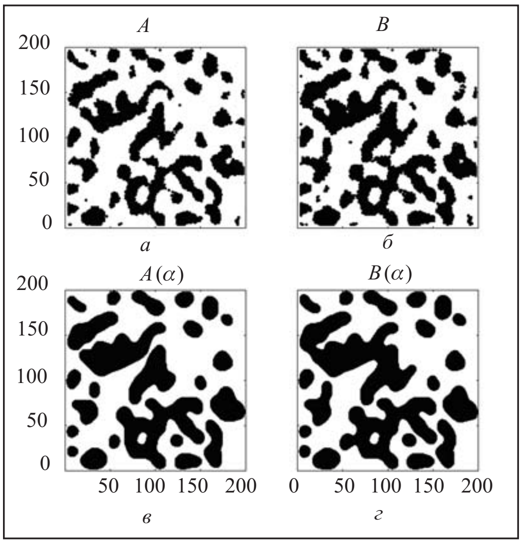

Рис. 1. Возрастающая зависимость алгоритма DPS по A

Замечание 1. Ранее отмечалось, что алгоритм DPS идемпотентен ($\text{DPS}^2 = \text{DPS}$). В совокупности с монотонностью по A это означает, что алгоритм DPS является морфологическим фильтром [11].

Утверждение 6. Если P, Q — плотности на X и $P_A(x) \leq Q_A(x)$ $\forall x \in X, A \subseteq X$, то $A_P(\alpha) \subseteq A_Q(\alpha)$.

Доказательство. Индукцией по n покажем, что $A_P^n(\alpha) \subseteq A_Q^n(\alpha)$.

Начало $n=1$: $x \in A_P^1(\alpha) \Leftrightarrow P_A(x) \ge \alpha \Rightarrow Q_A(x) \ge \alpha \Leftrightarrow x \in A_Q^1(\alpha)$.

Индукционный шаг $n-2 \to n-1$: предположение $A_P^{(n-1)}(\alpha) \subseteq A_Q^{(n-1)}(\alpha)$.

Доказательство перехода $n-1 \to n$: $x \in A_P^n(\alpha) \Leftrightarrow P_{A \cup A_P^{(n-1)}(\alpha)}(x) \ge \alpha$, следовательно, $P_{A \cup A_Q^{(n-1)}(\alpha)}(x) \ge \alpha \Rightarrow Q_{A \cup A_Q^{(n-1)}(\alpha)}(x) \ge \alpha \Leftrightarrow x \in A_Q^n(\alpha)$.

Таким образом, получаем две возрастающие подчиненные последовательности:

$$
\begin{array}{ccccccccc}
Q & & A_Q^1(\alpha) & \subseteq & A_Q^2(\alpha) & \subseteq & \ldots & = & A_Q^\infty(\alpha) \\
\nwarrow & \nearrow & & & & & & & \\
A & & \cup\!\mid & & \cup\!\mid & & & & \cup\!\mid \\
\swarrow & \searrow & & & & & & & \\
P & & A_P^1(\alpha) & \subseteq & A_P^2(\alpha) & \subseteq & \ldots & = & A_P^\infty(\alpha).
\end{array}
$$

Теперь индукцией по m покажем, что $A_P^{m\infty}(\alpha) \subseteq A_Q^{m\infty}(\alpha)$.

Начало $m=1$: выше доказано, что $A_P^\infty(\alpha) \subseteq A_Q^\infty(\alpha)$.

Индукционный шаг $m-2 \to m-1$: предположение $A_P^{(m-1)\infty}(\alpha) \subseteq A_Q^{(m-1)\infty}(\alpha)$.

Доказательство перехода $m-1 \to m$ основано на схеме (5):

$$A_P^{m\infty}(\alpha) = (A_P^{(m-1)\infty}(\alpha))^1 = \{x \in A_P^{(m-1)\infty}(\alpha) : P_{A_P^{(m-1)\infty}(\alpha)} \ge \alpha\}.$$

По предположению индукции $A_P^{(m-1)\infty}(\alpha) \subseteq A_Q^{(m-1)\infty}(\alpha)$, а вследствие монотонности P справедливо $P_{A_Q^{(m-1)\infty}(\alpha)} \ge P_{A_P^{(m-1)\infty}(\alpha)} \ge \alpha$. Подчиненность $P \le Q$ дает неравенство $Q_{A_Q^{(m-1)\infty}(\alpha)} \ge \alpha$, что равносильно включению $x \in A_Q^{m\infty}(\alpha)$.

Таким образом, получаем две неубывающие последовательности:

$$
\begin{array}{ccccccc}
A_Q^\infty(\alpha) & \supseteq & A_Q^{2\infty}(\alpha) & \supseteq & \ldots & = & A_Q(\alpha) \\
\cup\!\mid & & \cup\!\mid & & & & \cup\!\mid \\
A_P^\infty(\alpha) & \supseteq & A_P^{2\infty}(\alpha) & \supseteq & \ldots & = & A_P(\alpha),
\end{array}
$$

на концах которых будет нужное включение $A_P(\alpha) \subseteq A_Q(\alpha)$ (рис. 2).

Пример 4. Универсум X из примера 3, множество $A \subset X$ ($|A| = 10131$, см. рис. 2, *а*), плотности $P < Q$ имеют конструкцию $P_A(x) = |D_A(x, r)|$ с радиусами локализации $r_P = 2$, $r_Q = 3$, уровень $\alpha = 0.22$. В результате работы алгоритма DPS получены два множества: $A_P(\alpha) \subset A_Q(\alpha)$ ($|A_P(\alpha)| = 14069$, $|A_Q(\alpha)| = 30361$, см. рис. 2, *б*, *в*).

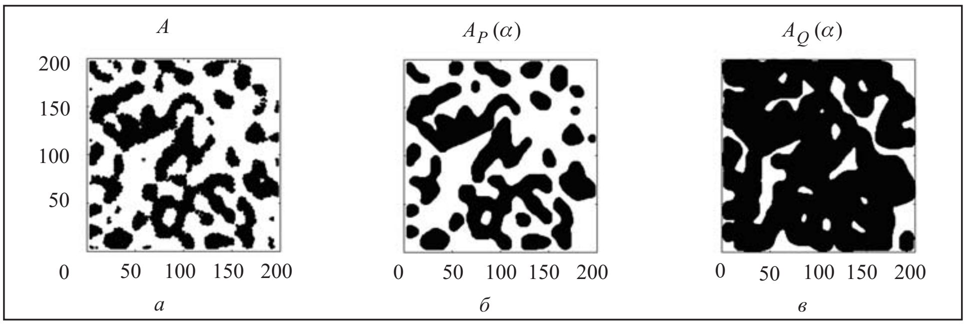

Рис. 2. Возрастающая зависимость алгоритма DPS по P

Утверждение 7. Если $\alpha < \beta$, то $A(\beta) \subseteq A(\alpha)$.

Доказательство. Далее с помощью индукции докажем включение $A^{\infty}(\beta) \subseteq A^{\infty}(\alpha)$. Считая, что оно установлено, согласно (3) и (1) имеем цепь импликаций

$$x \in A^{2\infty}(\beta) = (A^{\infty}(\beta))^1 \Rightarrow P_{A^{\infty}(\beta)}(x) \geq \beta \Rightarrow P_{A^{\infty}(\alpha)}(x) \geq \beta \Rightarrow P_{A^{\infty}(\alpha)}(x) \geq \alpha,$$

так что $x \in (A^{\infty}(\alpha))^1 = A^{2\infty}(\alpha)$ и поэтому $A^{2\infty}(\beta) \subseteq A^{2\infty}(\alpha)$. Далее для $m > 2$ аналогично получаем включения $A^{m\infty}(\beta) \subseteq A^{m\infty}(\alpha)$, откуда с учетом (6) $A(\beta) = \bigcap A^{m\infty}(\beta) \subseteq A^{m\infty}(\alpha) = A(\alpha)$.

Осталось индукцией доказать включение $A^{\infty}(\beta) \subseteq A^{\infty}(\alpha)$.

Начало $n = 1$: если $P_A(x) \geq \beta$, то $P_A(x) \geq \alpha \Rightarrow A^1(\beta) \subseteq A^1(\alpha)$.

Индукционный шаг $n-2 \to n-1$: предположение $A^{(n-1)}(\beta) \subseteq A^{(n-1)}(\alpha)$.

Доказательство перехода $n-1 \to n: x \in A^n(\beta) \Leftrightarrow P_{A \cup A^{(n-1)}(\beta)}(x) \geq \beta$, так что $P_{A \cup A^{(n-1)}(\beta)}(x) \geq \alpha \Rightarrow P_{A \cup A^{(n-1)}(\alpha)}(x) \geq \alpha \Leftrightarrow x \in A^n(\alpha)$ (последние переходы — следствия продолжения индукции и монотонности плотности P) и поэтому $A^{\infty}(\beta) = \bigcup A^n(\beta) \subseteq \bigcup A^n(\alpha) = A^{\infty}(\alpha)$.

Пример 5. Универсум X из примера 3, множество $A \subset X$ ($|A| = 10131$, см. рис. 3, *а*), плотность $P_A(x) = |D_A(x, r)|$ с радиусом лоализации $r = 3$, уровни $\alpha = 0.4$ и $\beta = 0.44$. В результате работы алгоритма DPS получены два множества: $A(\alpha) \supset A(\beta)$ ($|A(\alpha)| = 14747$, $|A(\beta)| = 10173$, см. рис. 3, *б*, *в*).

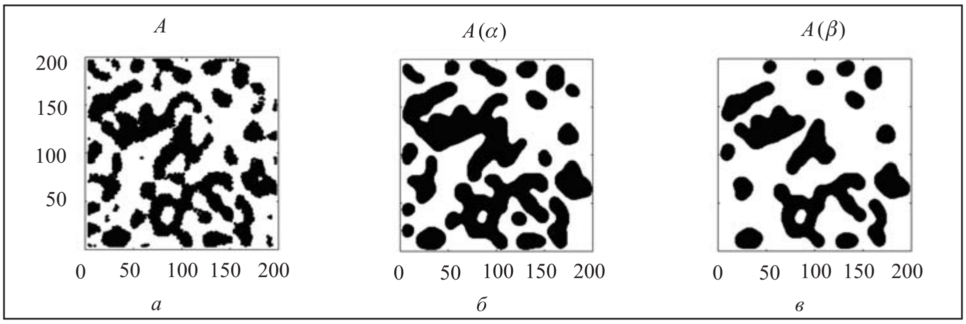

Рис. 3. Убывающая зависимость алгоритма DPS по α

Утверждение 8. Если $A \subseteq X \subset Y$ и мера P задана на Y, то $A_X(\alpha) \subseteq A_Y(\alpha)$.

Доказательство. Индукцией по n покажем, что $A_X^n(\alpha) \subseteq A_Y^n(\alpha)$.

Начало $n = 1$: $A_X^1(\alpha) = \{x \in X : P_A(x) \geq \alpha\} \subseteq \{y \in Y : P_A(y) \geq \alpha\} = A_Y^1(\alpha)$.

Индукционный шаг $n-2 \to n-1$: предположение $A_X^{(n-1)}(\alpha) \subseteq A_Y^{(n-1)}(\alpha)$.

Доказательство перехода $n-1 \to n: x \in A_X^n(\alpha) \Leftrightarrow P_{A \cup A_X^{(n-1)}(\alpha)}(x) \geq \alpha \Rightarrow$

$$\Rightarrow P_{A \cup A_Y^{(n-1)}(\alpha)}(x) \geq \alpha \Leftrightarrow x \in A_Y^n(\alpha).$$

Теперь индукцией по m покажем, что $A_X^{m\infty}(\alpha) \subseteq A_Y^{m\infty}(\alpha)$.

Начало $m = 1$ доказано выше: $A_X^{\infty}(\alpha) = \bigcup\limits_n A_X^n(\alpha) \subseteq \bigcup\limits_n A_Y^n(\alpha) = A_Y^{\infty}(\alpha)$.

Индукционный шаг $m-2 \to m-1$: предположение $A_X^{(m-1)\infty}(\alpha) \subseteq A_Y^{(m-1)\infty}(\alpha)$.

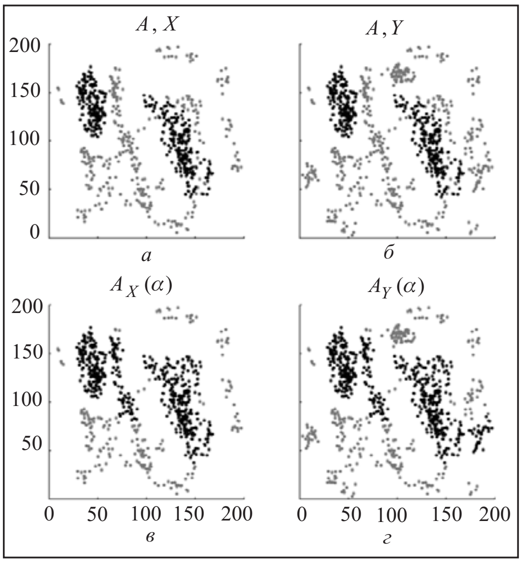

Рис. 4. Возрастающая зависимость алгоритма DPS по X

Доказательство перехода $m-1 \to m$ основано на схеме (5):

$$A_X^{m\infty}(\alpha) = (A_X^{(m-1)\infty}(\alpha))^1 =$$

$$= \{x \in A_X^{(m-1)\infty}(\alpha):$$

$$P_{A_X^{(m-1)\infty}(\alpha)}(x) \geq \alpha\} \subseteq$$

$$\subseteq \{y \in A_Y^{(m-1)\infty}(\alpha):$$

$$P_{A_Y^{(m-1)\infty}(\alpha)}(y) \geq \alpha\} = A_Y^{m\infty}(\alpha).$$

Последнее включение следует из предположения

$$A_X^{(m-1)\infty}(\alpha) \subseteq A_Y^{(m-1)\infty}(\alpha).$$

Пример 6. Заданы множества $A \subset X \subset Y$ ($|A| = 405$, $|X| = 729$, $|Y| = 865$, см. рис. 4, *а*, *б*). В обоих случаях радиус локализации $r = 14.35$ и уровень плотности $\alpha = 0.05$. В результате работы алгоритма DPS получены два множества: $A_X(\alpha) \subset A_Y(\alpha)$ ($|A_X(\alpha)| = 525$, $|A_Y(\alpha)| = 557$, см. рис. 4, *в*, *г*).

ВЫБОР ПАРАМЕТРОВ ДЛЯ АЛГОРИТМА DPS

Плотность. Плотность $P_A(x)$ при фиксированном A является нечеткой структурой на X. Поэтому с помощью операций нечеткой логики, а также некоторых других, можно получать новые плотности из существующих. Это расширяет возможности конструкции $A(\alpha)$ в изучении пространства X.

Утверждение 9. 1. Если P и Q — плотности на X и $R = R(y_1, y_2): [0,1] \times \times [0,1] \to [0,1]$ — неубывающее отображение, то $R(P,Q)_A(x) = R(P_A(x), Q_A(x))$ будет плотностью на X.

2. Если $\neg$ есть нечеткое отрицание [10], то

$$\neg P_A(x) = \neg(P_{\overline{A}}(x)) \tag{12}$$

будет плотностью на X.

3. Если φ — возрастающее отображение 2^X в 2^X, то $\varphi(P)_A(x) = P_{\varphi(A)}(x)$ является плотностью на X.

Доказательство. Проведем прямую проверку монотонности. Пусть $x \in X$ и $A \subset B \subseteq X$, тогда:

1) $P_A(X) \leq P_B(X)$ и $Q_A(X) \leq Q_B(X)$, следовательно,

$$R(P,Q)_A(x) = R(P_A(x), Q_A(x)) \leq R(P_B(x), Q_B(x)) = R(P,Q)_B(x);$$

2) $A \subset B \Rightarrow \overline{B} \subset \overline{A} \Rightarrow P_{\overline{B}}(x) \leq P_{\overline{A}}(x) \Rightarrow \neg(P_{\overline{B}}(x)) \geq \neg(P_{\overline{A}}(x))$, таким образом,

$$(\neg P)_A(x) = \neg(P_{\overline{A}}(x)) \leq \neg(P_{\overline{B}}(x)) = (\neg P)_B(x);$$

3) $A \subset B \Rightarrow \varphi(A) \leq \varphi(B)$, поэтому

$$\varphi(P)_A(x) = P_{\varphi(A)}(x) \leq P_{\varphi(B)}(x) = \varphi(P)_B(x).$$

Следствие 1. Плотностью будет R-соединение P и $\neg P$:

$$R(P, \neg P)_A(x) = R(P_A(x), \neg P_A(x)).$$

Следствие 2. 1. Если $\top$, $\perp$, M_w — соответственно t-норма, t-конорма, обобщенный оператор осреднения [10], то $\top(P_A(x), Q_A(x))$, $\perp(P_A(x), Q_A(x))$, $M_w(P_A(x), Q_A(x))$ являются плотностями на X.

2. Если $\lambda \in [0,1]$, то $\lambda P_A(x) + (1-\lambda)Q_A(x)$ и

$$\lambda P_A(x) + (1-\lambda)(1 - P_{\overline{A}}(x)) \tag{13}$$

являются плотностями на X.

Следствие 3. Для любого подмножества $C \subseteq X$ конструкции $P_A^{\cap C}(x) = P_{C \cap A}(x)$ и $P_A^{\cup C}(x) = P_{C \cup A}(x)$ являются плотностями на X.

В обозначениях следствия 3 анализ работы алгоритма DPS на множестве C относительно плотностей $P^{\cap C}$ и $P^{\cup C}$ позволит определить α-P-аналоги $\gamma_{P,\alpha}(C)$ и $\varphi_{P,\alpha}(C)$ морфологических операций открытия и закрытия на C [8, 11, 12].

Алгоритм DPS с плотностью $P^{\cap C}$. Для удобства объекты с тильдой будут элементами схемы (7), основанной на плотности $P^{\cap C}$:

$$\widetilde{C}^1 = \{x \in X : P_C^{\cap C}(x) \geq \alpha\} = \{x \in X : P_{C \cap C} \geq \alpha\} = C^1,$$

$$\widetilde{C}^2 = \{x \in X : P_{C \cup \widetilde{C}^1}^{\cap C}(x) \geq \alpha\} = \{x \in X : P_{C \cap (C \cup \widetilde{C}^1)} \geq \alpha\} = \{x \in X : P_C \geq \alpha\} = C^1,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\widetilde{C}^\infty = C^1.$$

Следовательно, в этой ситуации алгоритм DPS является убывающим. Стадия убывания имеет вид $\widetilde{C}^{m\infty} = \{x \in X : P_{C \cap \widetilde{C}^{(m-1)\infty}}(x) \geq \alpha\}$, и для полученного в конце α-$P^{\cap C}$-замыкания $\widetilde{C}(\alpha)$ справедливо соотношение $\widetilde{C}(\alpha) = \{x \in X : P_{\widetilde{C}(\alpha)}^{\cap C}(x) \geq \alpha\} = \{x \in X : P_{C \cap \widetilde{C}(\alpha)}(x) \geq \alpha\}$.

Утверждение 10. Пересечение $C \cap \widetilde{C}(\alpha)$ является α-P-совершенным в C и совпадает с $C_P(\alpha \mid C)$.

Доказательство. Покажем вначале, что $C \cap \widetilde{C}(\alpha) = \{x \in C : P_{C \cap \widetilde{C}(\alpha)}(x) \geq \alpha\}$.

Необходимость. Если $x \in C \cap \widetilde{C}(\alpha)$, то $x \in \widetilde{C}(\alpha) \Rightarrow P_{\widetilde{C}(\alpha)}^{\cap C}(x) = P_{C \cap \widetilde{C}(\alpha)}(x) \geq \alpha$.

Достаточность. Если $x \in X$ и $P_{C \cap \widetilde{C}(\alpha)}(x) \geq \alpha$, то $x \in \widetilde{C}(\alpha) \Rightarrow x \in C \cap \widetilde{C}(\alpha)$.

Теперь докажем равенство $C \cap \widetilde{C}(\alpha) = C_P(\alpha \mid C)$.

Необходимость. Множество $C_P(\alpha \mid C)$ — результат работы алгоритма DPS на C относительно P, и поэтому в силу монотонности DPS по множеству (утверждение 5) $C_P(\alpha \mid C)$ — максимальное α-P-совершенное множество в C. Следовательно, $C \cap \widetilde{C}(\alpha) \subseteq C_P(\alpha \mid C)$.

Достаточность. Плотность $P^{\cap C}$ на множестве 2^C подмножеств множества C совпадает с плотностью P, следовательно, в силу монотонности DPS по пространству (утверждение 8) $C_P(\alpha \mid C) = C_{P^C}(\alpha \mid C) \subseteq C_{P^C}(\alpha \mid X) = \widetilde{C}(\alpha)$ и поэтому $C_P(\alpha \mid C) = C_P(\alpha \mid C) \cap C \subseteq \widetilde{C}(\alpha) \cap C$.

Множество $C_P(\alpha \mid C)$ можно считать α-P-внутренностью C, а переход $\gamma_{\alpha,P}$ от C к $C_P(\alpha \mid C)$ — операцией α-P-открытия. Она удовлетворяет всем нужным для этого требованиям: $\gamma_{\alpha,P}(C) \subseteq C$, $\gamma_{\alpha,P}(C_1) \subseteq \gamma_{\alpha,P}(C_2)$, если $C_1 \subseteq C_2$ и $\gamma_{\alpha,P}^2 = \gamma_{\alpha,P}$.

Алгоритм DPS с плотностью $P^{\cup C}$. Как и ранее, объекты с тильдой являются элементами схемы (7), основанной на плотности $P^{\cup C}$:

$$\widetilde{C}^1 = \{x \in X : P_C^{\bigcup C}(x) \geq \alpha\} = \{x \in X : P_{C \bigcup C} \geq \alpha\} = C^1,$$

$$\widetilde{C}^2 = \{x \in X : P_{C \bigcup \widetilde{C}^1}^{\bigcup C}(x) \geq \alpha\} = \{x \in X : P_{C \bigcup (C \bigcup \widetilde{C}^1)} \geq \alpha\} =$$

$$= \{x \in X : P_{C \bigcup \widetilde{C}^1} \geq \alpha\} = C^2,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\widetilde{C}^n = C^n,\ n \geq 1 \Rightarrow \widetilde{C}^\infty = C^\infty.$$

Таким образом, для плотности $P^{\bigcup C}$ алгоритм DPS полностью повторяет возрастающую стадию с плотностью P. Рассмотрим стадию убывания

$$\widetilde{C}^{2\infty} = \{x \in X : P_{\widetilde{C}^{2\infty}}^{\bigcup C}(x) \geq \alpha\} = \{x \in X : P_{C \bigcup C^\infty}(x) \geq \alpha\} = C^\infty,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\widetilde{C}^{m\infty} = C^\infty \quad \forall m \geq 1.$$

Следовательно, алгоритм DPS на плотности $P^{\cup C}$ не имеет убывающей стадии, результатом его работы является множество C^∞ и оно α-$P^{\cup C}$-совершенно. Множество C^∞ ранее трактовалось как множество α-предельных для C точек. Объединение $C \bigcup C^\infty$ естественно считать α-P-замыканием C, а переход $\varphi_{\alpha,P}$ от C к $C \bigcup C^\infty$ — операцией α-P-замыкания. Она удовлетворяет всем нужным для этого требованиям: $\varphi_{\alpha,P}(C) \supseteq C$, $\varphi_{\alpha,P}(C_1) \subseteq \varphi_{\alpha,P}(C_2)$, если $C_1 \subset C_2$ и $\varphi_{\alpha,P}^2 = \varphi_{\alpha,P}$.

Остановимся подробнее на последнем свойстве. Из конструкции (7) следует, что

$$\varphi_{\alpha,P}(C)^1 = \{x \in X : P_{C \bigcup C^\infty}(x) \geq \alpha\} = C^\infty \subseteq \varphi_{\alpha,P}(C).$$

Так что $\varphi_P(C)^\infty = \varphi_P(C)^1 \subseteq \varphi_P(C)$ (утверждение 2) и поэтому

$$\varphi_{\alpha,P}(\varphi_{\alpha,P}(C)) = \varphi_{\alpha,P}(C) \bigcup \varphi_{\alpha,P}(C)^\infty = \varphi_{\alpha,P}(C).$$

Полученные выше результаты позволяют единообразно выразить операции $\gamma_{\alpha,P}$ и $\varphi_{\alpha,P}$:

$$\gamma_{\alpha,P}(C) = C \bigcap \mathrm{DPS}(C, X, P^{\bigcap C}, \alpha),$$

$$\varphi_{\alpha,P}(C) = C \bigcup \mathrm{DPS}(C, X, P^{\bigcup C}, \alpha).$$

Конкретные примеры плотностей рассматриваются, когда на множестве X задано расстояние d. Остановимся на двух вариантах определения плотности, для чего построим радиус локализации в X. Для этого обозначим $D(X)$ множество всех нетривиальных попарных расстояний в X:

$$D(X) = \{d(x, y) : x, y \in X, x \neq y\}.$$

Радиус локализации r для X определяется как степенное среднее всех расстояний из $D(X)$ при отрицательных значениях степени p:

$$r_p = r_p(X) = \left(\frac{\sum\limits_{d \in D(X)} d^p}{|D(X)|} \right)^{1/p}, \quad p < 0. \tag{14}$$

Первый вариант определения плотности «количество точек». Плотность зависит от радиуса локализации r (14) и параметра $q \geq 0$. Для каждой точки x из множества X рассматривается шар с центром в x радиуса r:

$$D(x, r) = \{ y \in X : d(x, y) \le r \}.$$

Для каждого шара вычисляется сумма точек множества X в нем с учетом расстояния от точек до его центра

$$N_X(x, r, q) = \sum_{y \in D(x, r)} \left(1 - \frac{d(x, y)}{r} \right)^q.$$

Максимум таких сумм по всем точкам $x \in X$ обозначим $C(X, r, q)$:

$$C(X, r, q) = \max_{x \in X} N_X(x, r, q).$$

Для каждого шара также вычисляется сумма точек с учетом их удаленности от центра шара только по точкам подмножества $A \subseteq X$:

$$N_A(x, r, q) = \sum_{y \in D_A(x, r)} \left(1 - \frac{d(x, y)}{r} \right)^q,$$

где $D_A(x, r)$ — пересечение шара и подмножества A: $D_A(x, r) = D(x, r) \bigcap A$.

При $q = 0$ получаем плотность «количество точек» в обычном смысле $P_A(x) = | D_A(x, r) |$, которая использовалась в примерах 3–5.

Плотность подмножества $A \subseteq X$ в точке $x \in X$ определяется как отношение суммы точек по шару в A с учетом их удаленности от центра к максимальной сумме по шарам в X:

$$P_A(x) = \frac{N_A(x, r, q)}{C(X, r, q)}.$$

Еще один способ перехода от функции $N_A(x, r, q)$ к плотности приведен далее (пример 7).

Второй вариант определения плотности «монолитность» [5]. Рассматривается проколотая окрестность точки x радиуса r: $D'_A(x, r) = \{ y \in A : 0 < d(x, y) \le r \}$. Она разбивается на m непересекающихся колец $D'_A(x, r) = \bigcup_{n=1}^{m} S_n$, где $S_n = \{ y \in A : r_{n-1} < d(x, y) \le r_n \}$, $0 = r_0 < \ldots < r_m = r$.

Каждому кольцу ставится в соответствие вес $1 \ge \psi_1 \ge \ldots \ge \psi_m > 0$.

Плотность подмножества $A \subseteq X$ в точке $x \in X$ определяется как отношение суммы весов непустых колец к сумме весов всех колец

$$P_A(x) = \frac{\sum_{n=1}^{m} \psi_n : S_n \ne \varnothing}{\sum_{n=1}^{m} \psi_n}.$$

Для проверки работы алгоритма DPS использовались веса, линейно убывающие при удалении от центра $\psi_n = 1 - \frac{n}{m+1}$, а также веса, полученные из них возведением в положительную степень $k > 0$: $\psi_n^k = \left(1 - \frac{n}{m+1} \right)^k$.

Уровень α. Выбор уровня α сильно влияет на результат работы алгоритма DPS. Удобным инструментом для выбора уровня α являются нечеткие сравнения [12]. Нечеткое сравнение $n(a, b)$ двух положительных чисел: a и b, — мера превосходства числа b над числом a со значениями на отрезке $[-1, 1]$: $n(a, b) = \text{mes}\,(a < b) \in [-1, 1]$.

Нечеткое сравнение числа a и конечного множества чисел B можно определить как среднее нечетких сравнений a со всеми числами из множества B:

$$n(a, B) = \frac{\sum_{b \in B} n(a, b)}{|B|}, \quad n(B, a) = \frac{\sum_{b \in B} n(b, a)}{|B|},$$

и понимать как меру минимальности в случае $n(a, B)$ и меру максимальности в случае $n(B, a)$ числа a на фоне B.

При фиксированном B функции $n(a, B)$ и $n(B, a)$ монотонны и непрерывны по a. Таким образом, по результату сравнения a с B однозначно определяется непосредственно a.

Пример 7. Положим $N_A(r, q) = \{N_A(x, r, q) : x \in X\}$. Мера сравнения $N_A(r, q)$ со значениями $N_A(x, r, q)$ позволяет получить новую плотность $P_A(x) = (n(\text{Im}\, N_A(r, q), N_A(x, r, q)) + 1)/2$, где n — нечеткое сравнение [13]:

$$n(a, b) = \frac{b - a}{\max(|a|, |b|)}.$$

Мера максимальности $n(B, a)$ дает возможность сформулировать необходимое требование к будущему результату $A \subset X$ алгоритма DPS: плотность множества A в каждой своей точке x должна быть значительной (достаточно максимальной) на фоне X. Для этого вначале рассчитаем значения плотности всего универсума X во всех его точках: $P_X(X) = \{P_X(x) : x \in X\}$, т.е. «фон X». Если $\beta \in [-1, 1]$ — необходимый уровень максимальности плотности P на фоне X, то непосредственный уровень $\alpha = \alpha(\beta)$ для P однозначно определяется по β из уравнения:

$$n(P_X(X), \alpha) = \beta. \qquad (15)$$

В силу монотонности n искать такое α можно в виде решения уравнения (15) методом деления отрезка пополам.

Таким образом, с помощью алгоритма DPS можно найти в X подмножество A, которое β-экстремально P-плотное в каждой своей точке x: $n(P_X(X), P_A(x)) \geq \beta \Leftrightarrow P_A(x) \geq \alpha(\beta)\ \forall x \in X$.

Пример 8. Рассмотрим результат работы убывающего алгоритма DPS на всем пространстве X при различных уровнях плотности β. Радиус локализации $r = 5.35$ получен при $p = -3$ (14). Повышая уровень β, можно проникнуть вглубь сгущений пространства X, тем самым изучая их морфологию. На рис. 5, *а* показан результат работы алгоритма DPS при уровне плотности $\beta = -0.45$, а на рис. 5, *б* — при уровне плотности $\beta = -0.1$.

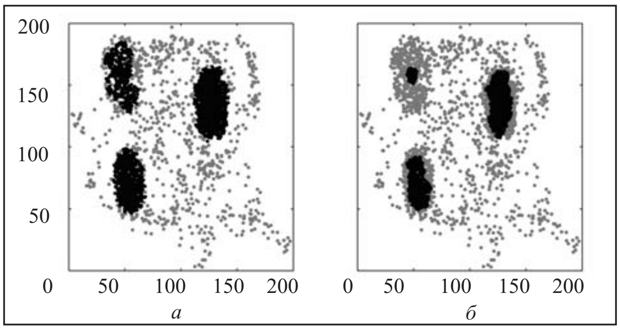

Рис. 5. Результаты работы алгоритма DPS на синтетическом примере

Замечание 2. Проиллюстрируем на пространстве X (рис. 6) разницу между простым выбором по уровню плотности и результатом работы DPS. Результат простого выбора (см. рис. 6, *а*) — множество $X^1(\alpha)=\{x\in X: P_X(x)\ \ge\ \alpha\}$ $(\alpha=\alpha(\beta)$, $\beta=-0.45)$, не является α-плотным во всех своих точках, в то время как результат работы DPS (см. рис. 6, *б*) — множество $A(\alpha)$, α-плотно во всех своих точках. Результат работы алгоритма DPS не содержит отдельных изолированных точек в отличие от простого выбора по уровню плотности.

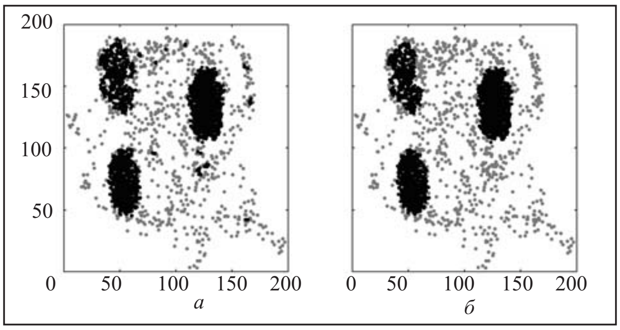

Рис. 6. Результаты кластеризации алгоритмом DPS на синтетическом примере с простым выбором по уровню плотности

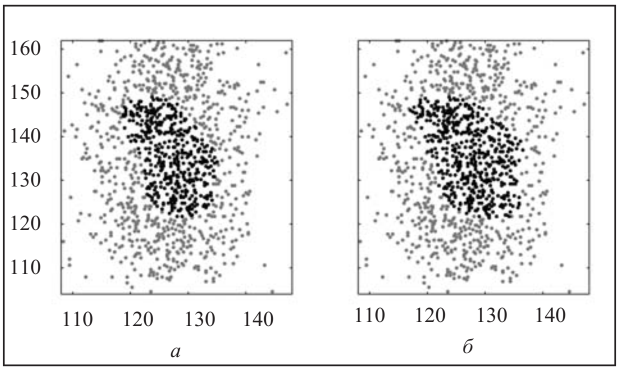

Рис. 7. Результаты работы алгоритма DPS при различных λ

Замечание 3. Если алгоритм DPS с плотностью P настроен на поиск P-плотных множеств, то DPS с плотностью $\neg P$ (12) — скорее, на поиск P-отделимых множеств. Их λ-соединение (13) позволяет одновременно учесть и плотность, и отделимость в работе DPS. На рис. 7, *а* показаны результаты работы алгоритма DPS при $\lambda=0$, а на рис. 7, *б* — при $\lambda=0.35$. Множество, полученное во втором случае, визуально более отделимо, чем множество, полученное в первом.

Критерий качества связан со свойствами (9) и (10) нечеткой структуры P_A на X для α-совершенного множества A: $\forall x\in A$, $y\in\overline{A}$ $P_A(x)\ge\alpha>P_{\overline{A}}(y)$ или $P_A(A)=\{P_A(x), x\in A\}\ge\alpha>$ $>\{P_A(y), y\in\overline{A}\}=P_A(\overline{A})$.

Качеством $\tau(A)$ множества A считается всякое формальное выражение преимущества числовой совокупности $P_A(A)$ над числовой совокупностью $P_A(\overline{A})$.

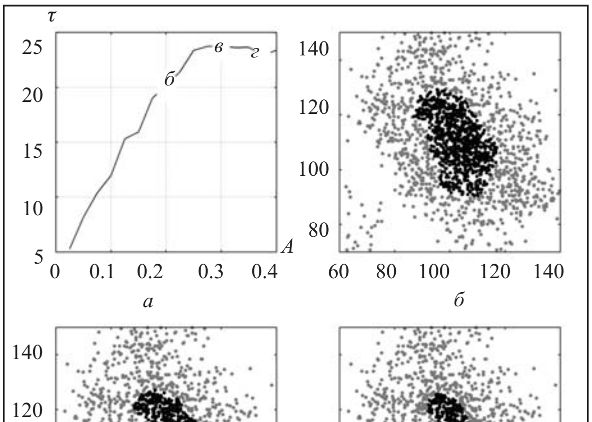

Рис. 8. График зависимости $\tau(A)$ и результаты работы DPS при различных $\tau(A)$

Замечание 4. Исследования показывают, что конструкция степенного усреднения $M_P(B)$ неотрицательного множества B с отрицательным показателем $k\le-2$ выполняет функцию статистически устойчивого минимума B, а с показателем $q\ge 2$ — статистически устойчивого максимума B. На основании этого можно построить первую версию качества множества A:

$$\tau(A)=\tau_{kq}(A)=M_k(P_A(A))-M_q(P_A(\overline{A})),\ k<0,\ q>0.$$

Пример 9. При $k=-\infty$, $q=\infty$ имеет место $\tau(A)=\min P_A(A)-\max P_A(\overline{A})$.

Пример 10. На рис. 8, *а* показан график зависимости $\tau(A)=M_{-2}(P_A(A))-$ $-M_2(P_A(\overline{A}))$ от значения уровня α, на рис. 8, *б*, *в*, *г* — результаты работы DPS при различных $\tau(A)$. Наилучший результат достигается при максимальном значении критерия качества (см. рис. 8, *в*).

Пример 11. Второй критерий является мерой превосходства $P_A(A)$ над $P_A(\overline{A})$ на основе нечеткого сравнений *n*:

$$\tau(A)=n(P_A(\overline{A}),P_A(A))=\frac{\sum\limits_{x\in A,\ y\in\overline{A}} n(P_A(y),P_A(x))}{|A||\overline{A}|}.$$

Качество τ позволяет оптимизировать выбор параметра α в работе алгоритма DPS на множестве A: $\alpha^*=\alpha^*(A)=\arg\min\tau(\mathrm{DPS}(A\,|\,X,P,\alpha))$.

НЕЧЕТКАЯ ИНТЕРПРЕТАЦИЯ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

В процессе анализа конструкции $A(\alpha)$ установлена обратная зависимость по α всех элементов схемы (7), что дает возможность связать с каждым из них нечеткие структуры на X. В стандартных обозначениях теории нечетких множеств [10, 14] положим для $n,m\geq 1$:

$$\mathbf{A}^n=\bigcup_\alpha A^n(\alpha),\ \mathbf{A}^n(x)=\sup\{\alpha\in[0,1]:x\in A^n(\alpha)\},$$

$$\mathbf{A}^{m\infty}=\bigcup_\alpha A^{m\infty}(\alpha),\ \mathbf{A}^{m\infty}(x)=\sup\{\alpha\in[0,1]:x\in A^{m\infty}(\alpha)\},\qquad(16)$$

$$\mathbf{DPS}A=\bigcup_\alpha A(\alpha),\ \mathbf{DPS}A(x)=\sup\{\alpha\in[0,1]:x\in A(\alpha)\}.$$

Соответствия (16) сохраняют порядок на нечетких структурах. Кроме того, при любом $\alpha\in[0,1]$ как возрастающая стадия $A^n(\alpha)\uparrow A^\infty(\alpha)$, так и убывающая $A^{m\infty}(\alpha)\downarrow A(\alpha)$ ограничены по длине порядком множества X. Поэтому возрастающий ряд нечетких структур $\mathbf{A}^n$ заканчивается структурой $\mathbf{A}^\infty$, а убывающий ряд $\mathbf{A}^{m\infty}$ — структурой **DPS** *A*. Все это позволяет получить нечеткую интерпретацию процесса построения $A\to A(\alpha)$ (7), «проинтегрировав его по α»:

$$A\to\mathbf{A}^1\subseteq\ldots\subseteq\mathbf{A}^\infty\supseteq\ldots\supseteq\mathbf{DPS}\,A.$$

Плотность $P_A(x)$ можно интерпретировать как меру принадлежности x к A в X. В этом смысле P_A является первой и самой естественной структурой, описывающей влияние множества A на пространство X через плотность P. В приведенных обозначениях P_A совпадает с $\mathbf{A}^1$ и не всегда согласована со структурой **DPS** *A*, выражающей влияние A на X через DPS.

Тем не менее, в случае $A=X$ структуры P_X и $\mathbf{DPS}X$ согласованы.

Утверждение 11. Имеет место неравенство $\mathbf{DPS}\,X\leq P_X$.

Доказательство. Если $x\in X$ и $\alpha\in[0,1]$, то из определения **DPS**, совершенности $X(\alpha)$ и монотонности P следует цепочка неравенств

$$\mathbf{DPS}\,X(x)\geq\alpha\Leftrightarrow\{x\in X(\alpha)\}\Leftrightarrow P_{X(\alpha)}(x)\geq\alpha\Rightarrow P_X(x)\geq\alpha,$$

а поэтому $\mathbf{DPS}\,A(x)\leq P_X(x)$.

Очевидно, **DPS**-влияние A на X является плотностью, поскольку конструкция $A(\alpha)$ положительно монотонна по A (см. утверждение 5). Положим $\mathbf{DPS}_A(x) = \mathbf{DPS}\ A(x)$ и исследуем плотность **DPS**. Прямые вычисления показывают, что первая оболочка $A^1_{\mathbf{DPS}}(\alpha)$ любого множества A относительно плотности **DPS** совпадает с его α-совершенным замыканием $A(\alpha) = A_P(\alpha)$ относительно плотности P:

$$A^1_{\mathbf{DPS}}(\alpha) = \{x \in X : \mathbf{DPS}_A(x) \geq \alpha\} \Leftrightarrow \{x \in X : x \in A(\alpha)\} = A_P(\alpha). \qquad (17)$$

Отсюда, а также из утверждения 4 ($A^1 = A$) следует наследственность **DPS**-плотности по отношению к P.

Утверждение 12. Всякое α-P-совершенное множество α-**DPS**-совершенно.

Доказательство. Имеем

$$\left.\begin{array}{c} A = A_P(\alpha) \\ A^1_{\mathbf{DPS}}(\alpha) = A_P(\alpha) \end{array}\right\| \Rightarrow A = A^1_{\mathbf{DPS}}(\alpha) \Leftrightarrow A \text{ согласно (11) } \alpha\text{-}\mathbf{DPS}\text{-совершенно.}$$

Соотношение (17) позволяет понять работу алгоритма DPS относительно «своей плотности»:

$$A^1_{\mathbf{DPS}}(\alpha) = A_P(\alpha),$$

$$A^2_{\mathbf{DPS}}(\alpha) = (A + A^1_{\mathbf{DPS}}(\alpha))^1_{\mathbf{DPS}} = (A + A^1_{\mathbf{DPS}}(\alpha))(\alpha),$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$A^{i+1}_{\mathbf{DPS}}(\alpha) = (A + A^i_{\mathbf{DPS}}(\alpha))^1_{\mathbf{DPS}} = (A + A^i_{\mathbf{DPS}}(\alpha))(\alpha).$$

Стабилизация $A^{\infty}_{\mathbf{DPS}}(\alpha)$ будет α-P-совершенной, а поэтому в силу наследственности и α-**DPS**-совершенной. Таким образом, имеем условия возрастающего **DPS** и $A_{\mathbf{DPS}}(\alpha) = A^{\infty}_{\mathbf{DPS}}(\alpha)$.

ЗАКЛЮЧЕНИЕ

Созданный в ГЦ РАН дискретный математический анализ (ДМА) — новый подход к изучению многомерных массивов и временных рядов, базирующийся на нечеткой математике и нечеткой логике, представляет собой серию алгоритмов, предназначенных для решения основных задач анализа данных: кластеризации и трассирования в многомерных массивах, изучения временных рядов по функциональному сценарию, подобному классическому анализу гладких функций.

Все алгоритмы ДМА универсальны и скреплены единой формальной основой, в которой фундаментальные понятия предельности, непрерывности, плотности, близости, связности и тренда выражены в рамках нечеткой математики и нечеткой логики. В свою очередь, из них конструируются все алгоритмы ДМА для обработки данных.

Кластерная составляющая ДМА основана на формальном понятии плотности, которая позволяет строго определить для многомерного массива понятия сгущения (плотное на общем фоне подмножество), кластеры (изолированное сгущение) и трассы (локально линейное сгущение). В ДМА их поиск осуществляется с помощью алгоритмов Роден, Кристалл и Монолит, а также представленного в данной статье алгоритма DPS. Они предназначены для топологической фильтрации массива, позволяющей отсекать от него несущественные (неплотные) части.

В настоящее время алгоритмы ДМА-кластеризации эффективно примененяются в геофизических науках (интерпретация магнитных и гравитационных полей, спутниковая интерферометрия), экологии (поиск мест возможного захоронения ядерных отходов). На результатах работы алгоритм DPS основан анализ сейсмических каталогов для распознавания мест возможного возникновения сильных землетрясений [15, 16].

СПИСОК ЛИТЕРАТУРЫ

1. Алгоритмы искусственного интеллекта для кластеризации магнитных аномалий / А.Д. Гвишиани, М. Диаман, В.О. Михайлов и др. // Физика Земли. — 2002. — № 7. — С. 13–28.
2. Гвишиани А.Д., Агаян С.М., Богоутдинов Ш.Р. О новом подходе к кластеризации // Кибернетика и системный анализ. — 2002. — № 2. — С. 104–122.
3. Application of artificial intelligence for Euler solutions clustering / V. Mikhailov, A. Galdeano, M. Diament et al. // Geophysics. — 2003. — **68**, N 1. — P. 168–180.
4. Агаян С.М., Соловьев А.А. Выделение плотных областей в метрических пространствах на основе кристаллизации // System Res. & Inform. Technol. — 2004. — N 2. — P. 7–23.
5. Гвишиани А.Д., Агаян С.М., Богоутдинов Ш.Р., Соловьев А.А. Дискретный математический анализ и геолого-геофизические приложения // Вестн. КРАУНЦ. Науки о Земле. — 2010. — № 2, вып. 16. — С. 109–125.
6. Геоинформационные технологии: методы искусственного интеллекта при оценке тектонической стабильности Нижнеканского массива / А.Д. Гвишиани, С.В. Белов, С.М. Агаян и др. // Инженерная экология. — 2008. — № 2. — С. 3–14.
7. Everitt B.S. Cluster analysis. — London: Halsted-Heinemann, 1980. — 170 p.
8. Ту Дж., Гонсалес Д. Принципы распознавания образов. — М.: Мир, 1978. — 412 с.
9. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Классификация и снижение размерности: Справ. изд. / Под ред. С.А. Айвазяна. — М.: Финансы и статистика, 1989. — 606 с.
10. Нечеткие множества в моделях управления и искусственного интеллекта / А.Н. Аверкин, И.З. Батыршин, А.Ф. Блишун и др. / Под ред. Д.А. Поспелова. — М.: Наука, 1986. — 312 с.
11. Serra J. Image analysis and mathematical morphology. — New York: Acad. press, 1982. — 621 p.
12. Гонсалес Р., Вудс Р. Цифровая обработка изображений. — М.: Техносфера, 2005. — 1071 с.
13. Математические методы геоинформатики. III. Нечеткие сравнения и распознавание аномалий на временных рядах / А.Д. Гвишиани, С.М. Агаян, Ш.Р. Богоутдинов и др. // Кибернетика и системный анализ. — 2008. — **44**, № 3. — С. 3–18.
14. Орловский С.А. Проблемы принятия решений при нечеткой исходной информации. — М.: Наука, 1981. — 208 с.
15. Gvishiani A., Dobrovolsky M., Agayan S., Dzeboev B. Fuzzy-based clustering of epicenters and strong earthquake-prone areas // Environ. Eng. and Manag. J. — 2013. — **12**, N 1. — P. 1–10.
16. Гвишиани А.Д., Агаян С.М., Добровольский М.Н., Дзебоев Б.А. Объективная классификация эпицентров и распознавание мест возможного возникновения сильных землетрясений в Калифорнии // Геоинформатика. — 2013. — № 2. — С. 44–57.

Поступила 22.02.2013