

УДК 519.68:007.5

ВЫДЕЛЕНИЕ ПЛОТНЫХ ОБЛАСТЕЙ В МЕТРИЧЕСКИХ ПРОСТРАНСТВАХ НА ОСНОВЕ КРИСТАЛЛИЗАЦИИ

С.М. АГАЯН, А.А. СОЛОВЬЕВ

Рассмотрено применение методов искусственного интеллекта и теории нечетких множеств к задаче кластеризации. Объект исследования — данные о плотных скоплениях в конечных метрических пространствах. В работе использован математический аппарат на базе теории нечетких множеств и «оптический» подход к анализу метрических пространств. Результат исследования — реализация серии алгоритмов «Кристалл» поиска плотных скоплений в многомерных массивах данных.

ВВЕДЕНИЕ

Статья продолжает цикл работ «Математические методы геоинформатики» [1–5], посвященных применению в геофизике методов нечеткой теории множеств и нечеткой логики. Подход с точки зрения нечеткой математики используется для выделения плотных областей в конечных метрических пространствах. Работа выполнена при поддержке Российского фонда фундаментальных исследований (грант 04-05-65055-а).

Опыт работы с геофизическими массивами данных показывает, что большое значение имеют области повышенной плотности. В двумерном массиве X на рис. 1 они выделены специальным образом. Изолированные сгущения являются предметом кластерного анализа [6]. Нас интересует нахождение их в произвольном случае.

В работах [1–3] используется высекание из X слабейших, наименее плотных точек (алгоритм «Роден»). В настоящей статье предлагается дуальная стратегия присоединения сильнейших, наиболее плотных точек в X , своего рода кристаллизация в X (алгоритм «Кристалл»). Неформальная суть ее состоит в следующем: в качестве основы кристалла выбираются наиболее плотные точки в X (на рис.1 они отмечены звездочками). Возьмем одну из них и обозначим x^* . Затем устроим борьбу между x^* и ее дополнением $X - x^*$ за право обладания другой точкой $y \in X - x^*$. Если основа кристалла x^* победила в некоторых точках из дополнения $X - x^*$, то к ней присоединяется та из них, в которой эта победа была особенно убедительной. Обозначим ее x_1 . Так получается рост кристалла K от его начальной версии K_0 к следующей версии K_1 :

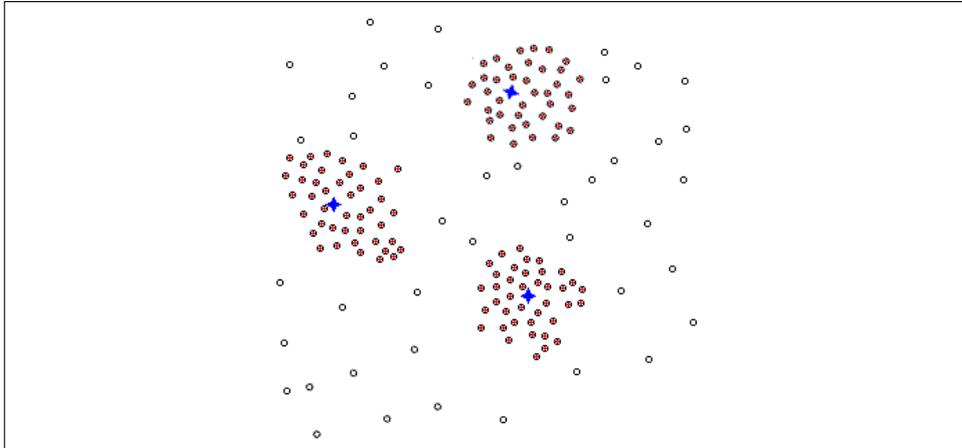


Рис. 1. Пример сгущений в двумерном пространстве

$$K_0 = \{x_0 = x^*\} \rightarrow K_1 = \{x_0, x_1\}.$$

Далее за свои точки с первой версией кристалла K_1 борется уже его дополнение $X - K_1$. И опять рост кристалла K произойдет путем присоединения к K_1 точки x_2 , в которой победа K_1 над $\overline{K_1} = X - K_1$ будет наиболее убедительной:

$$K_1 \rightarrow K_2 = \{K_1, x_2\}.$$

Если дополнение $\overline{K_1}$ сохранило в борьбе с первой версией K_1 кристалла K все свои точки, то рост K считается законченным, и кристалл K полагается равным K_1 : $K = K_1$. Это и будет наиболее плотной в X частью, содержащей x^* . На следующих стадиях действует та же логика: кристалл K растет в своей n -й версии K_n , если в $\overline{K_n}$ существуют точки, в которых K_n побеждает $\overline{K_n}$, и прекращает свой рост, останавливаясь на K_n в противном случае.

Формализация кристаллизации будет выполнена на основе оптического подхода [1–3]: введенный в конечном метрическом пространстве X свет $\delta_x(y)$ позволит придать точный смысл всем встречавшимся выше понятиям: плотность, борьба и победа.

1. ОПТИЧЕСКИЙ ПОДХОД

Пусть (X, d) — конечное метрическое пространство.

1.1. Свет

Закон распространения света $\delta_x(\cdot)$ от источника в точке $x \in X$ является свободным параметром изложения и имеет различную математическую конструкцию. Функция $\delta_x(y)$ моделирует близость в X точки y к точке x и

может считаться нечеткой мерой принадлежности y к x . Таким образом, в общем случае $\delta_x(y)$ не только убывает с ростом расстояния $d(x, y)$, но и зависит от топологии X вокруг x .

Приведем несколько примеров.

1.1.1. С каждой убывающей на положительной полуоси $[0, \infty)$ неотрицательной функцией «потенциального» типа $\varphi: \varphi(0) = 1, \varphi(t_1) \geq \varphi(t_2), t_1 < t_2$ связан свет $\delta_x^\varphi(y) = \delta_x(y)$.

$$\delta_x(y) = \varphi(d(x, y)), \quad \forall y \in X.$$

1.1.2. Обычная эксклюзивная максимальность: именно такие $\bar{y} \in X$ говорят в пользу близости y к x в X .

$$\delta_x(y) = \frac{|\bar{y} \in X : d(x, \bar{y}) > d(x, y)| + 1}{|X|}.$$

Обозначим $X(x) = U_{i=1}^{n(x)} C_x(r_i)$ — центрированное представление X вокруг x , где $C_x(r_i) \stackrel{\text{def}}{=} \{y \in X : d(x, y) = r_i\}$ — окружности радиусов r_i , на которых располагаются точки в X относительно x ; $\Gamma(x) = \{0 < r_1 < \dots < r_{n(x)}\}$ — «группа нормирования» X в x . Перейдем от $X(x)$ к наиболее его адекватному одномерному выражению, а именно неотрицательному распределению масс на прямой: $\Gamma X(x) = \{(r_i, m_i = |C_x(r_i)|) \}_1^{n(x)}$.

Приводимые ниже меры $\delta_x(y)$ связаны с движением (рис.2) слева направо, начиная от нуля, расстояния $d(x, y)$ сквозь распределение $\Gamma X(x)$: $\delta_x(y)$ — степень непроникновения $d(x, y)$ в $\Gamma X(x)$ ($\equiv 1$ — степень глубины $d(x, y)$ в $\Gamma X(x)$). Разное моделирование движения приводит к разным мерам.

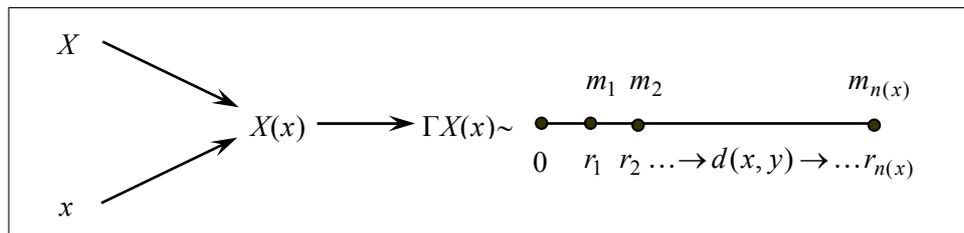


Рис. 2. Пример мер: движение в $\Gamma X(x)$

Нам понадобится нечеткое сравнение $n: \square^+ \times \square^+ \rightarrow [0, 1]$ на неотрицательных числах $a, b \geq 0$: $n(a, b)$ — мера превосходства b над a со свойствами

$$n(0, b) \equiv 1, \quad \forall b,$$

$$n(a, 0) \equiv 0, \quad \forall a,$$

$$n(a,b) > 1/2 \Leftrightarrow a < b,$$

$$n(a,b) < 1/2 \Leftrightarrow a > b,$$

$$n(a,b) = 1/2 \Leftrightarrow a = b.$$

В качестве n возьмем

$$n(a,b) = \frac{b - a + \max(a,b)}{2 \max(a,b)}.$$

1.1.3. Движение в $\Gamma X(x)$ через моменты $M^+(y)$ и $M^-(y)$. Положим для $y \in X$

$$M^+(y) = (\sum (d(x,y) - r_i) m_i : r_i < d(x,y)),$$

$$M^-(y) = (\sum (r_i - d(x,y)) m_i : r_i > d(x,y)).$$

Моменты $M^+(y)$, $M^-(y)$ можно считать доводом в пользу максимальной (минимальной) $d(x,y)$ в $\Gamma X(x)$. Их нечеткое сравнение даст нужную меру

$$\delta_x(y) = n(M^-(y), M^+(y)).$$

1.1.4. Движение в $\Gamma X(x)$ через сравнение с радиусами. Положим для $y \in X$

$$\begin{aligned} \delta_x(y) &= \frac{\sum_{i=1}^{n(x)} m_i n(d(x,y), r_i)}{\sum_{i=1}^{n(x)} m_i} = \\ &= \frac{\sum_{i=1}^{n(x)} m_i n(d(x,y), r_i)}{|X|-1} = \frac{\sum_{\bar{y} \neq x} n(d(x,y), d(\bar{y}, y))}{|X|-1}. \end{aligned}$$

1.2. Плотность освещения

Для точки $x \in X$ и произвольного подмножества $A \subset X$ положим

$$A_x \stackrel{\text{def}}{=} \begin{cases} A, & \text{если } x \notin A, \\ A - x, & \text{если } x \in A, \end{cases} \quad |A|_x \stackrel{\text{def}}{=} |A|_x = \begin{cases} |A|, & \text{если } x \notin A, \\ |A|-1, & \text{если } x \in A. \end{cases}$$

1.2.1. **Определение 1.** Назовем плотностью освещения $P_A(x)$ точки x подмножеством A средний свет, приходящий в x от источников $y \in A$

$$P_A(x) = \frac{\sum_{y \in A_x} \delta_y(x)}{|A|_x}.$$

Определение 2. Назовем плотностью A и обозначим

$$P(A) = \min_{x \in A} P_A(x) .$$

Результаты исследований подтверждают способность плотности численно выражать интегральную сконцентрированность A вокруг x . Продолжая аналогию с нечеткой принадлежностью, можно считать $P_A(x)$ нечеткой мерой принадлежности x к A . Оказывается, что не только зависимость $x \rightarrow P_A(x)$ плотности по первой компоненте обладает хорошими свойствами, но и вторая зависимость $A \rightarrow P_A(x)$. Это имеет фундаментальное значение [1].

1.2.2. **Утверждение.** Зависимость $A \rightarrow P_A(x)$ линейна: если $A, B \subset X$ и $A \cap B = \emptyset$, то

$$P_{A \cup B}(x) = \frac{|A|_x}{|A|_x + |B|_x} P_A(x) + \frac{|B|_x}{|A|_x + |B|_x} P_B(x) .$$

В частности, при дополнительном предположении $x \in A$

$$P_{A \cup B}(x) = \frac{|A| - 1}{|A| + |B| - 1} P_A(x) + \frac{|B|}{|A| + |B| - 1} P_B(x) .$$

Сформулируем предмет нашего поиска в X .

1.2.3. **Определение.** Подмножество $A \subset X$ плотно на фоне X , если

$$P_A(x) \geq P_x(x), \quad \forall x \in A .$$

Искать такие A в X будем глобальным вариантом алгоритма «Кристалл».

2. Глобальный «Кристалл»

Параметры алгоритма: конечное метрическое пространство (X, d) , модель света $\delta_x(\cdot)$, параметр роста кристалла α , параметр основы β .

2.1. Блок-схема (рис. 3)

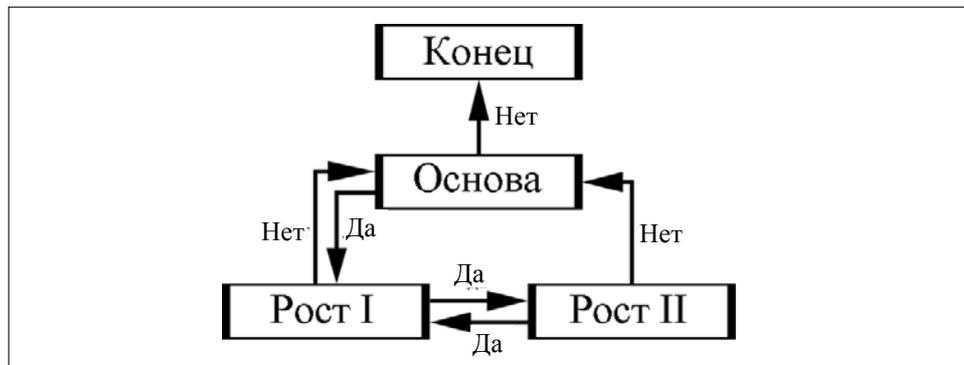


Рис. 3. Блок-схема алгоритма

2.2. Основа

Задача блока состоит в выборе начальных точек (основ) кристаллизации. Последние должны быть достаточно плотны в X . В глобальном «Кристалле» множество основ $F \subset X$ определяется с помощью глобальной функции плотности: если $\Sigma P(X)$ — среднее значение $P_x(\cdot)$ на X , $\beta \geq 1$, то

$$F \stackrel{\text{def}}{=} F_\beta = \{x \in X : P_x(x) \geq \beta \Sigma P(X)\}.$$

Самая плотная точка x^* в X — начало первого кристалла $K^{(1)}$.

$$K_0^{(1)} = x^* = x_0 = \arg \max_x P_x(\cdot).$$

Если закончена i -я кристаллизация $K^{(i)}$ и при этом $F \subset K^1 \cup \dots \cup K^i$, то алгоритм завершает свою работу (переход в блок «Конец») и $K^{(1)}, \dots, K^{(i)}$ — его результат. В противном случае $F \not\subset K^1 \cup \dots \cup K^i$ происходит выбор $x_0^{(i+1)}$ — основы следующей $(i+1)$ -й кристаллизации.

$$x_0^{(i+1)} = \arg \max_{F \setminus (K^1 \cup \dots \cup K^i)} P_x(\cdot).$$

Далее осуществляется переход в блок «Рост I».

2.3. Рост I

Если K_n^i — текущая версия i -го кристалла K^i , то борьба K_n^i с \bar{K}_n^i в точке $x \in \bar{K}_n^i$ формализуется в виде частного $\frac{P_{K_n^i}(x)}{P_{\bar{K}_n^i}(x)}$, а результат борьбы — в виде сравнения этого частного с порогом роста α .

Если $P_{K_n^i}(x) < P_{\bar{K}_n^i}(x)$, то \bar{K}_n^i сохранило за собой свою точку x . Рост кристалла K^i заканчивается на стадии K_n^i , если его дополнение \bar{K}_n^i сохранило все свои точки. Далее происходит переход к блоку «Основа».

Если $P_{K_n^i}(x) \geq P_{\bar{K}_n^i}(x)$, то K_n^i одержал над \bar{K}_n^i победу в точке x , но кристалл K^i может перейти из K_n^i в следующую стадию K_{n+1}^i через x : $K_{n+1}^i = \{K_n^i, x\}$ только при условии сохранения своей плотности на фоне X . И с этим связана вторая проверка роста кристалла.

2.4. Рост II

Вывод условия на x , гарантирующего сохранение плотности $K_{n+1}^i = \{K_n^i, x\}$ на фоне X , содержит два утверждения.

2.4.1 Утверждение 1. n -я версия K_n кристалла K плотна на фоне $X \Leftrightarrow P_{K_n}(x) \geq P_{\bar{K}_n}(x)$, $\forall x \in K_n$.

Доказательство. Из представления $X = K_n \cup \bar{K}_n$ и линейности плотности 1.2.2 следует для $x \in K_n$ равенство

$$P_X(x) = \frac{|K_n|-1}{|X|-1} P_{K_n}(x) + \frac{|\bar{K}_n|}{|X|-1} P_{\bar{K}_n}(x).$$

Поэтому условие плотности 1.2.3 примет вид

$$P_{K_n}(x) \geq \frac{|K_n|-1}{|X|-1} P_{K_n}(x) + \frac{|\bar{K}_n|}{|X|-1} P_{\bar{K}_n}(x).$$

Принимая во внимание равенство $|K_n| + |\bar{K}_n| = |X|$, имеем

$$P_{K_n}(x) \geq \frac{|K_n|-1}{|X|-1} P_{K_n}(x) + \frac{|X|-|K_n|}{|X|-1} P_{\bar{K}_n}(x)$$

или

$$P_{K_n}(x) \geq P_{\bar{K}_n}(x) \quad \forall x \in K_n.$$

Ч.т.д.

2.4.2. Следствие. Порог роста кристалла $\alpha \geq 1$.

Доказательство. Если кристалл K из стадии K_n переходит в стадию $K_{n+1} = \{K_n, x_{n+1}\}$, то необходимо, согласно 2.4.1, неравенство $P_{K_{n+1}}(x_{n+1}) \geq P_{\bar{K}_{n+1}}(x_{n+1})$, но $P_{K_{n+1}}(x_{n+1}) = P_{K_n}(x_{n+1})$ и $P_{\bar{K}_{n+1}}(x_{n+1}) = P_{\bar{K}_n}(x_{n+1})$, а потому $\alpha \geq 1$.

Ч. т. д.

Итак, если K_n победил \bar{K}_n в точке $x_{n+1} \in \bar{K}_n$, то плотность $K_{n+1} = \{K_n, x_{n+1}\}$ на фоне X в x_{n+1} имеет место автоматически ($\alpha \geq 1$). Осталось проверить условие 2.4.1 для K_{n+1} в остальных его точках, т. е. для $x \in K_n$. Нам потребуется выразить $P_{\bar{K}_{n+1}}(x)$ через $P_{\bar{K}_n}(x)$ и $\delta_{x_{n+1}}(x)$: $\bar{K}_n = \bar{K}_{n+1} \cup x_{n+1}$, $x \in K_n$, следовательно (линейность 1.2.2)

$$P_{\bar{K}_n}(x) = \frac{|\bar{K}_{n+1}|}{|\bar{K}_n|} P_{\bar{K}_{n+1}}(x) + \frac{1}{|\bar{K}_n|} \delta_{x_{n+1}}(x) = \frac{|\bar{K}_n|-1}{|\bar{K}_n|} P_{\bar{K}_{n+1}}(x) + \frac{1}{|\bar{K}_n|} \delta_{x_{n+1}}(x)$$

и

$$P_{\bar{K}_{n+1}}(x) = \frac{|\bar{K}_n|}{|\bar{K}_n|-1} P_{\bar{K}_n}(x) - \frac{1}{|\bar{K}_n|-1} \delta_{x_{n+1}}(x).$$

Равенства

$$P_{K_{n+1}}(x) = \frac{|K_n|-1}{|K_n|} P_{K_n}(x) + \frac{1}{|K_n|} \delta_{x_{n+1}}(x), \quad |K_n| + |\bar{K}_n| = |X|$$

и

$$P_X(x) = \frac{|K_n|-1}{|X|-1} P_{K_n}(x) + \frac{|\bar{K}_n|}{|X|-1} P_{K_n}(x) \Leftrightarrow$$

$$\Leftrightarrow |\bar{K}_n| P_{\bar{K}_n}(x) = (|X|-1) P_X(x) - (|K_n|-1) P_{K_n}(x)$$

дают возможность выразить неравенство $P_{K_{n+1}}(x) \geq P_{\bar{K}_{n+1}}(x)$, $\forall x \in K_n$ через n -ю стадию кристалла

$$\frac{|K_n|-1}{|K_n|} P_{K_n}(x) + \frac{1}{|K_n|} \delta_{x_{n+1}}(x) \geq \frac{|\bar{K}_n|}{|\bar{K}_n|-1} P_{\bar{K}_n}(x) - \frac{1}{|\bar{K}_n|-1} \delta_{x_{n+1}}(x) \Leftrightarrow$$

$$\Leftrightarrow \left(\frac{1}{|K_n|} + \frac{1}{|\bar{K}_n|-1} \right) \delta_{x_{n+1}}(x) \geq \frac{|\bar{K}_n|}{|\bar{K}_n|-1} P_{\bar{K}_n}(x) - \frac{|K_n|-1}{|K_n|} P_{K_n}(x) =$$

$$= \frac{(|X|-1) P_X(x) - (|K_n|-1) P_{K_n}(x)}{|\bar{K}_n|-1} - \frac{(|K_n|-1) P_{K_n}(x)}{|K_n|} = \frac{(|X|-1)}{|\bar{K}_n|-1} P_X(x) -$$

$$- (|K_n|-1) \left(\frac{1}{|\bar{K}_n|-1} + \frac{1}{|K_n|} \right) P_{K_n}(x) \Leftrightarrow \frac{|X|-1}{|K_n| (|\bar{K}_n|-1)} \delta_{x_{n+1}}(x) \geq \frac{|X|-1}{|\bar{K}_n|-1} P_X(x) -$$

$$- \frac{(|X|-1)(|K_n|-1)}{|K_n| (|\bar{K}_n|-1)} P_{K_n}(x) \Leftrightarrow \delta_{x_{n+1}}(x) \geq |K_n| P_X(x) - (|K_n|-1) P_{K_n}(x).$$

Сформулируем утверждение, являющееся сутью блока «Рост II».

2.4.3. Утверждение 2. Рост кристалла $K_n \rightarrow K_{n+1} = K_n \cup x_{n+1}$ через точку x_{n+1} описывается неравенствами

$$P_{K_n}(x) \geq P_{\bar{K}_n}(x_{n+1}) \quad (*)$$

и

$$\delta_{x_{n+1}}(x) \geq |K_n| P_X(x) - (|K_n|-1) P_{K_n}(x), \quad \forall x \in K_n. \quad (**)$$

Итак, в блоке «Рост II» происходит проверка точек x_{n+1} (*), которые K_n отрывает от \bar{K}_n , на сохранение плотности с их участием. Если они не выдерживают условие (**), то текущий кристалл расти перестает и в алгоритме происходит переход к блоку «Основа» с целью поиска основы для следующей кристаллизации. Если таких x_{n+1} оказывается несколько $(x_{n+1}^1, \dots, x_{n+1}^{m_n})$, то в качестве продолжения берется та из них, в которой качество кристалла K на $(n+1)$ -й стадии будет максимальным (1.2.1).

$$x_{n+1} = \arg \max_{s=1}^{m_n} P(k_n \cup x_{n+1}^s).$$

Далее происходит расчет плотностей $P_{K_{n+1}}(\cdot)$ и $P_{\bar{K}_{n+1}}(\cdot)$ на \bar{K}_{n+1} для борьбы между ними за рост кристалла K на следующем $(n+1)$ -м уровне в блоке «Рост I».

$$P_{K_{n+1}}(x) = \frac{|K_n|}{|K_n|+1} P_{K_n}(x) + \frac{1}{|K_n|+1} \delta_{x_{n+1}}(x), \quad \forall x \in \bar{K}_{n+1},$$

$$P_{\bar{K}_{n+1}}(x) = \frac{|\bar{K}_n|-1}{|\bar{K}_n|-2} P_{\bar{K}_n}(x) - \frac{1}{|\bar{K}_n|-2} \delta_{x_{n+1}}(x), \quad \forall x \in \bar{K}_{n+1}.$$

2.5. Конец

В данном блоке происходит идентификация окончательных версий кристалла K^1, \dots, K^I , $I = I(F, \alpha)$.

3. ЛОКАЛЬНЫЙ «КРИСТАЛЛ»

Рассмотрим ситуацию, изображенную на рис. 4. Процесс кристаллизации K начнется где-то в середине A и в общем случае будет непременно происходить то слева, то справа от K_0 . Через какое-то время из-за больших линейных размеров A освещение еще не принадлежащих K справа точек из A правым краем A и множеством B может стать более плотным, чем освещение этих точек кристаллом K . В результате кристалл K прекратит свой рост, не достигнув правого конца A . Хотя из рис. 4 ясно видно, что именно A должен быть окончательной его версией: $K = A$.

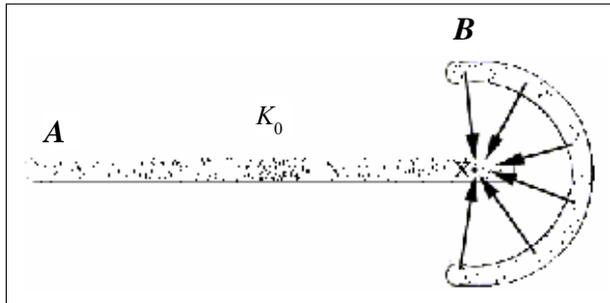


Рис. 4. Пример необходимости применения локального «Кристалла»

Выйти из этого положения можно, освободившись от большой длины растущего кристалла с помощью так называемой локальной кристаллизации, рассмотрев борьбу между K и \bar{K} в точке x не во всем пространстве X , а только в его шаре определенного радиуса r с центром в x .

3.1. Локальная плотность

Для ее определения потребуется несколько обозначений: $D(x, r) = \{y \in X : d(x, y) \leq r\}$ — шар в X с центром в x и радиусом r . Если $A \subset X$, то $D_A(x, r) = D(x, r) \cap A = \{y \in A : d(x, y) \leq r\}$ — часть в A , отстоящая от x не более, чем на r . Заметим, что $D_A(x, r)$ может быть пусто и, кроме того, $x \notin A \Rightarrow x \notin D_A(x, r)$.

С заданной моделью света $\delta_x(\cdot)$ на X можно связать не только плотность $P_A(\cdot)$, но и целую серию производных от нее так называемых локальных r -плотностей $P_A(\cdot)(r)$.

3.1.1. Определение 1. Назовем локальной r -плотностью подмножества A в точке x

$$P_A(x)(r) \stackrel{\text{def}}{=} P_{D_A(x,r)}(x) = \begin{cases} \frac{1}{|D_A(x,r)|_x} \sum_{\substack{y \in D_A(x,r) \\ y \neq x}} \delta_y(x), & \text{если } |D_A(x,r)|_x > 0, \\ 0, & \text{если } |D_A(x,r)|_x = 0. \end{cases}$$

Определение 2. Назовем локальной r -плотностью A и обозначим

$$P(A)(r) = \min_{x \in A} P_A(x)(r).$$

Таким образом, r -плотность A в x равна нулю в двух случаях: когда x лежит в A , но при этом r -изолирован в нем ($D_A(x,r) = x$), или когда x r -отделим от A ($D_A(x,r) = \emptyset$).

Далее, переход $A \rightarrow D_A(x,r)$ согласован с дизъюнктым объединением

$$\begin{aligned} D_{A \vee B}(x,r) &= D(x,r) \cap (A \vee B) = \\ &= (D(x,r) \cap A) \vee (D(x,r) \cap B) = D_A(x,r) \vee D_B(x,r). \end{aligned}$$

Отсюда следует линейность (1.2.2) локальной плотности

$$P_{A \vee B}(x)(r) = \frac{|D_A(x,r)|_x}{|D_{A \vee B}(x,r)|_x} P_A(x)(r) + \frac{|D_B(x,r)|_x}{|D_{A \vee B}(x,r)|_x} P_B(x)(r).$$

3.1.2. Определение. Подмножество $A \subset X$ r -плотно на фоне X , если

$$P_A(x)(r) \geq P_X(x)(r), \quad \forall x \in A.$$

Цель локального r -«Кристалла» — поиск r -плотных A в X . Его параметры — пространство (X, d) , модель света $\delta_x(\cdot)$, радиус r , параметр роста кристалла α , параметр основы β .

Изложение r -«Кристалла» аналогично изложению глобального «Кристалла» и будет вестись по его модулю.

3.2. Основа

Если $\sum P(X)(r)$ — среднее значение плотности $P_X(\cdot)(r)$ на X ; $F = F(\beta, r)$ — множество основ локальной r -кристаллизации, то

$$F \stackrel{\text{def}}{=} \left\{ x \in X : P_X(x) \geq \beta \sum P(X)(r) \right\}.$$

Самая r -плотная точка x^* в X — начало первого r -кристалла K^1 .

$$K_0^1 = x^* = x_0 = \arg \max_X P_X(\cdot)(r).$$

Остальное аналогично 2.2.

3.3. Рост I

Для удобства положим

$$D_n(x, r) = D_{K_n}(x, r),$$

$$D_{\bar{n}}(x, r) = D_{\bar{K}_n}(x, r),$$

$$P_n(x, r) = P_{K_n}(x, r),$$

$$P_{\bar{n}}(x, r) = P_{\bar{K}_n}(x, r).$$

При этом $D(x, r) = D_n(x, r) \vee D_{\bar{n}}(x, r)$.

Если K_n — текущая версия кристалла K , то ее борьба с \bar{K}_n в точке $x \in \bar{K}_n$ происходит в шаре $D(x, r)$ между $D_n(x, r)$ и $D_{\bar{n}}(x, r)$. Как и в глобальном «Кристалле», результатом борьбы является сравнение частного $\frac{P_n(x)(r)}{P_{\bar{n}}(x)(r)}$ с порогом α . Особенность локального случая состоит в том, что локальные плотности $P_n(x)(r)$ и $P_{\bar{n}}(x)(r)$ могут равняться нулю. Рассмотрим все случаи.

3.3.1. $P_{\bar{n}}(x)(r) > 0$ и $P_n(x)(r) < \alpha P_{\bar{n}}(x)(r)$: x остается за \bar{K}_n . Случай $P_n(x)(r) > 0$ приведен на рис. 5, а $P_n(x)(r) = 0$ — на рис. 6.

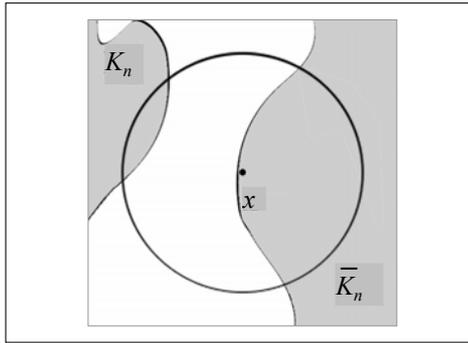


Рис. 5. Случай $P_n(x)(r) < \alpha P_{\bar{n}}(x)(r)$, $P_n(x)(r) > 0$

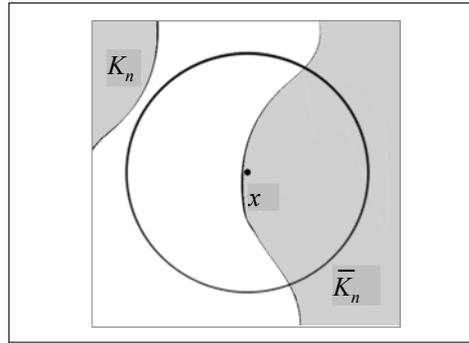


Рис. 6. Случай $P_n(x)(r) < \alpha P_{\bar{n}}(x)(r)$, $P_n(x)(r) = 0$

3.3.2. $P_n(x)(r) \geq \alpha P_{\bar{n}}(x)(r)$: K_n отрывает x у \bar{K}_n , но на следующем этапе в блоке «Рост II» может отвергнуть x в качестве точки роста. Случай $P_{\bar{n}}(x)(r) > 0$ изображен на рис. 7, а $P_{\bar{n}}(x)(r) = 0$ — на рис. 8.

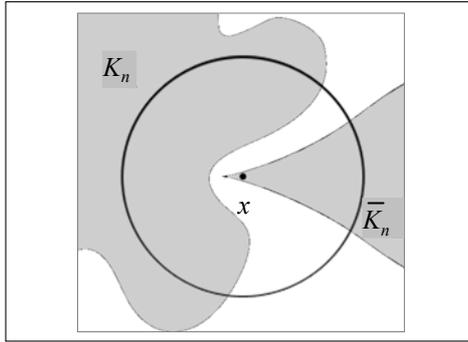


Рис. 7. Случай $P_n(x)(r) \geq \alpha P_{\bar{n}}(x)(r)$, $P_{\bar{n}}(x)(r) > 0$

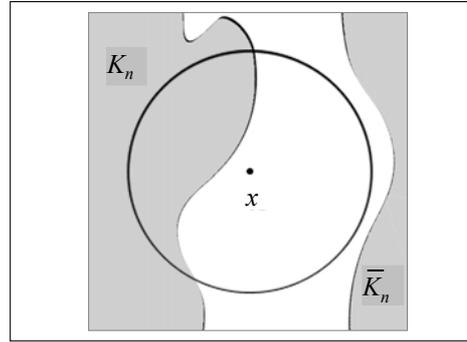


Рис. 8. Случай $P_n(x)(r) \geq \alpha P_{\bar{n}}(x)(r)$, $P_{\bar{n}}(x)(r) = 0$

3.3.3. $P_n(x)(r) = P_{\bar{n}}(x)(r) = 0$: точка x абсолютно r -изолирована в X , не может служить продолжением для K_n и потому остается в \bar{K}_n (рис.9).

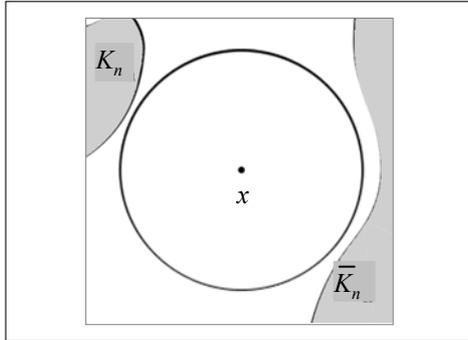


Рис. 9. Случай $P_n(x)(r) = P_{\bar{n}}(x)(r) = 0$

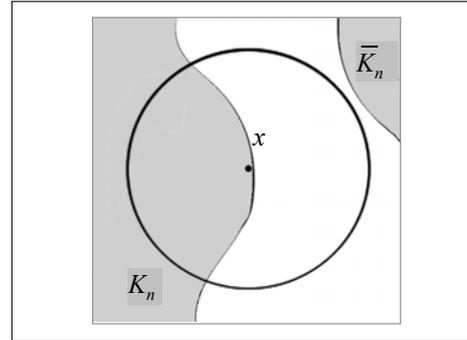


Рис. 10. Случай $P_X(x)(r) = P_n(x)(r) \times (|D_{\bar{n}}(x,r)| = 0)$

3.4. Рост II

По аналогии с глобальным случаем условие на x , гарантирующее сохранение локальной плотности $K_{n+1} = K_n \vee x_{n+1}$ на фоне X , заключено в двух утверждениях:

3.4.1. **Утверждение 1.** n -я версия K_n кристалла K локально r -плотна на фоне $X \Leftrightarrow P_n(x)(r) \geq P_{\bar{n}}(x)(r)$, $\forall x \in K_n$.

Доказательство. Из представления $D(x,r) = D_n(x,r) \vee D_{\bar{n}}(x,r)$ и локальной линейности $P_x(\cdot)(r)$ для x справедливо

$$P_X(x)(r) = \frac{|D_n(x,r)|_x}{|D(x,r)|_x} P_n(x)(r) + \frac{|D_{\bar{n}}(x,r)|_x}{|D(x,r)|_x} P_{\bar{n}}(x)(r).$$

Но $x \in D_n(x,r) \Rightarrow |D_n(x,r)|_x = |D_n(x,r)| - 1$, $|D_{\bar{n}}(x,r)|_x = |D_{\bar{n}}(x,r)|$. Если $|D_{\bar{n}}(x,r)| = 0$, то $D_n(x,r) = D(x,r)$ и $P_X(x)(r) = P_n(x)(r)$ (рис.10). Если $|D_{\bar{n}}(x,r)|_x > 0$, то аналогично 2.4.1 $P_n(x)(r) \geq P_{\bar{n}}(x)(r)$ и $\alpha \geq 1$ (рис. 11).

Ч. г. д.

Это означает, что, как и в глобальном «Кристалле», плотность кристалла K_{n+1} в точке x_{n+1} , поступившей из блока «Рост I», выполнена автоматически. В силу локальности, проверку 3.4.1 в остальных $x \in K_{n+1}$ нужно делать только при условии принадлежности $x_{n+1} \in D(x, r) \Leftrightarrow x \in D(x_{n+1}, r)$ (рис. 12).

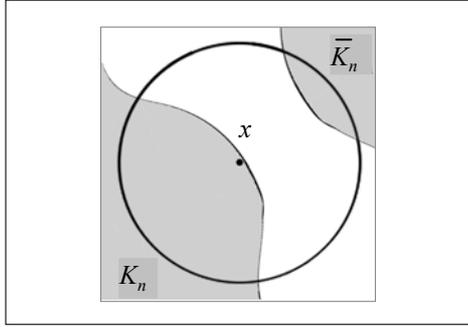


Рис. 11. Случай $P_n(x)(r) \geq P_{\bar{n}}(x)(r) \times (|D_{\bar{n}}(x, r)| > 0)$

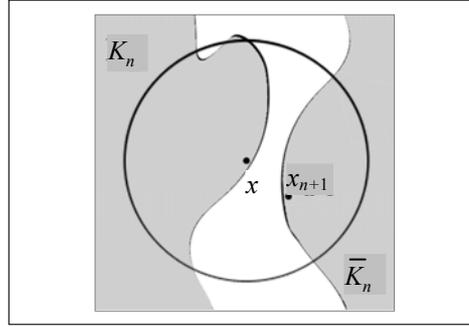


Рис. 12. Случай $x_{n+1} \in D(x, r) \times (x \in D(x_{n+1}, r))$

Заменяя в 2.4.3 X на $D(x, r)$, K_n на $D_n(x, r)$ и \bar{K}_n на $D_{\bar{n}}(x, r)$, получаем локальный критерий плотности продолжения для x_{n+1} .

3.4.2. **Утверждение 2.** Пусть $x_{n+1} \in K_{\bar{n}}$ и $P_n(x_{n+1})(r) \geq \alpha P_{\bar{n}}(x_{n+1})(r)$. Переход $K_n \rightarrow K_{n+1} = K_n \vee x_{n+1}$ с сохранением локальной плотности на фоне X возможен только при условии

$$\delta_{x_{n+1}}(x) \geq |D_n(x, r)| P_X(x)(r) - (|D_n(x, r)| - 1) P_n(x)(r), \quad \forall x \in D_n(x_{n+1}, r).$$

Если $x_{n+1}^1, \dots, x_{n+1}^{m_n}$ — возможные продолжения, то выбор среди них осуществляется по качеству (3.1.1).

$$x_{n+1} = \arg \max_{s=1}^{m_n} P(K_n + x_{n+1}^s)(r).$$

Далее происходит расчет плотностей $P_{n+1}(\cdot)(r)$ и $P_{\bar{n}+1}(\cdot)(r)$ на \bar{K}_{n+1} для борьбы между ними за рост кристалла K на следующем $(n+1)$ -м уровне в блоке «Рост I»: если $x \in \bar{K}_{n+1}$ и $d(x, x_{n+1}) > r$, то ее не затронули предыдущие события и все в ней осталось без изменений.

$$D_{n+1}(x, r) = D_n(x, r), \quad D_{\bar{n}+1}(x, r) = D_{\bar{n}}(x, r),$$

$$P_{n+1}(x)(r) = P_n(x)(r), \quad P_{\bar{n}+1}(x)(r) = P_{\bar{n}}(x)(r).$$

Если $d(x, x_{n+1}) \leq r$, то

$$D_{n+1}(x, r) = D_n(x, r) + x_{n+1}, \quad D_{\bar{n}+1}(x, r) = D_{\bar{n}}(x, r) - x_{n+1}$$

и

$$P_{n+1}(x)(r) = \frac{|D_n(x,r)|}{|D_n(x,r)|+1} P_n(x)(r) + \frac{1}{|D_n(x,r)|+1} \delta_{x_{n+1}}(x),$$

$$P_{n+1}^-(x)(r) =$$

$$= \begin{cases} \frac{|D_n^-(x,r)|-1}{|D_n^-(x,r)|-2} P_n^-(x)(r) - \frac{1}{|D_n^-(x,r)|-2} \delta_{x_{n+1}}(x), & \text{если } |D_n^-(x,r)| > 2, \\ 0, & \text{если } |D_n^-(x,r)| \leq 2. \end{cases}$$

Второй случай означает, что после перехода x_{n+1} из K_n^- в K_{n+1} точка x стала в K_{n+1}^- r -изолированной.

3.5. Конец

В данном блоке происходит идентификация окончательных версий кристалла K^1, \dots, K^I , $I = I(F, \alpha, r)$.

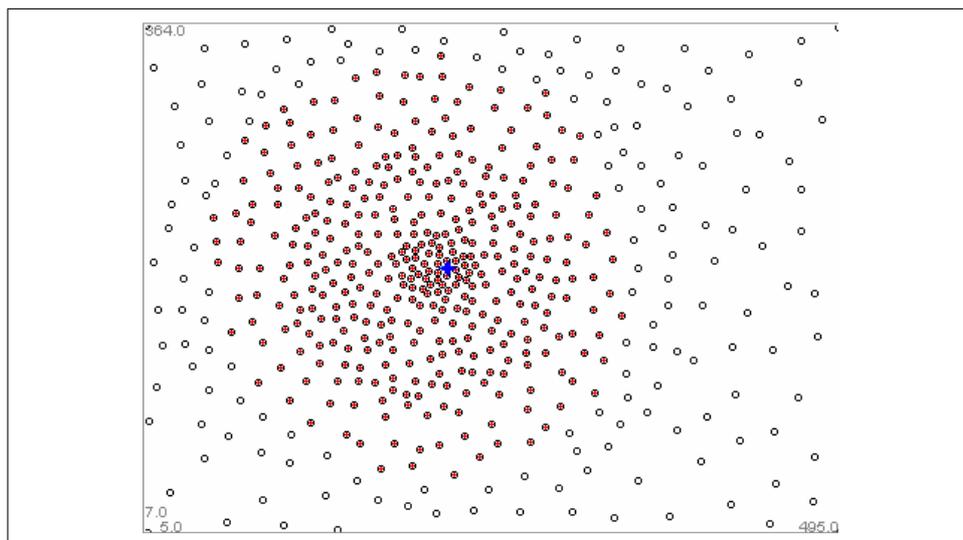


Рис. 13. Пример кристаллизации вокруг одной основы, $\gamma = 1,0$

Завершая описание «Кристалла», сделаем несколько замечаний.

1. В отличие от «Родена», «Кристалл» сразу имеет дело с плотным на фоне X множеством.
2. В плотном на фоне X подмножестве A не все точки обязаны быть сильными и годятся на роль основ из F : кристаллизация дает в отличие от F более целостную картину уплотнений в X .
3. Параметры α и β отвечают за жесткость «Кристалла»: большим α и β соответствуют более компактные и сильные уплотнения в X .
4. В блоке «Основа» подмножество F может быть получено с помощью другой плотности, построенной на иных принципах.

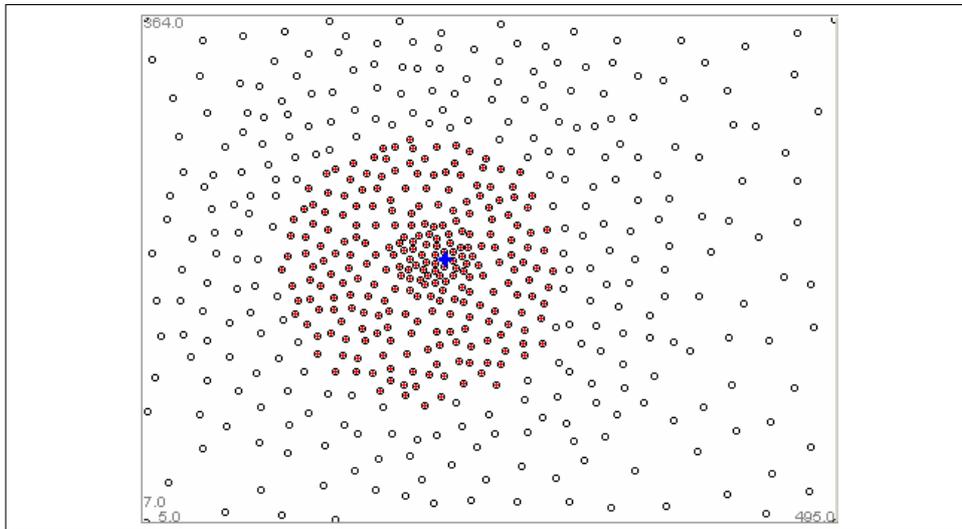


Рис. 14. Пример кристаллизации вокруг одной основы, $\gamma = 1,2$

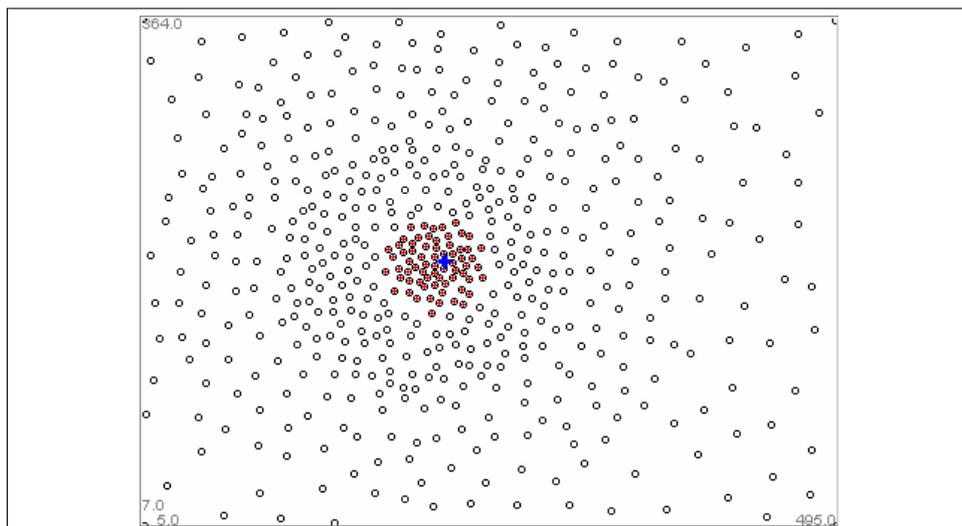


Рис. 15. Пример кристаллизации вокруг одной основы, $\gamma = 2,3$

5. Неспособность K_n оторвать у \bar{K}_n точки равносильно неравенству $P_{\bar{K}_n}(x) > P_{K_n}(x)$, $\forall x \in \bar{K}_n$. Согласно 2.4.1, это означает повышенную плотность \bar{K}_n на фоне X . Так в процессе кристаллизации K_n мы можем получить уплотнения с обратной стороны.

6. В локальном «Кристалле» радиусы могут быть переменными и зависеть от x : $r = r(x)$. Кроме того, мягкими в «Кристалле» могут быть неравенства, связанные с α и β : в обозначениях 1.1.2 $a < b$, если $n(a,b) > 1/2$ или $3/4$.
мягко

7. Исследования показали, что довольно часто требуется γ -усиление плотности на общем фоне, а именно γ -аналог 1.2.3 при $\gamma > 1$.

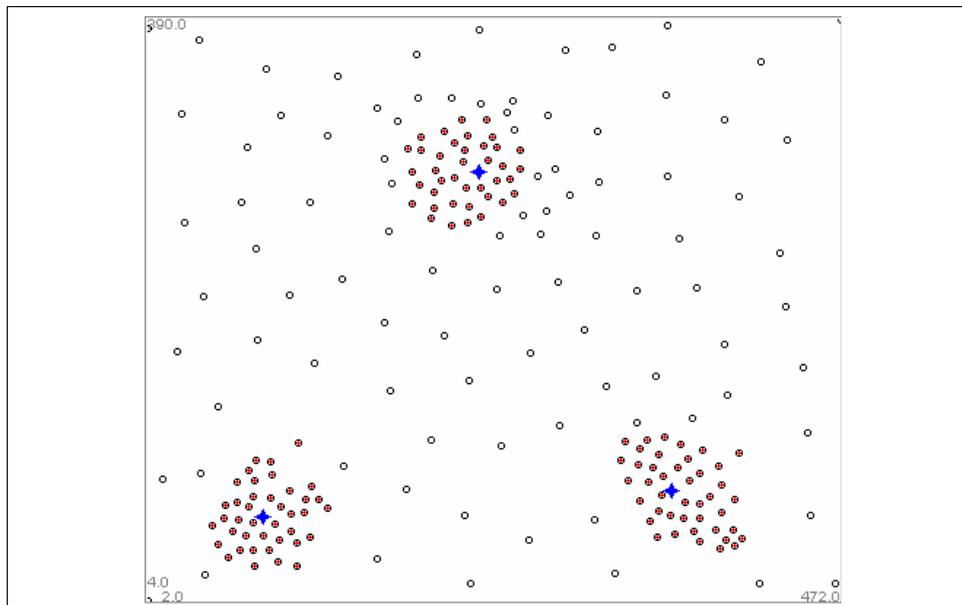


Рис. 16. Результат работы «Кристалла», $\gamma = 2,5$; $\beta = 1,2$

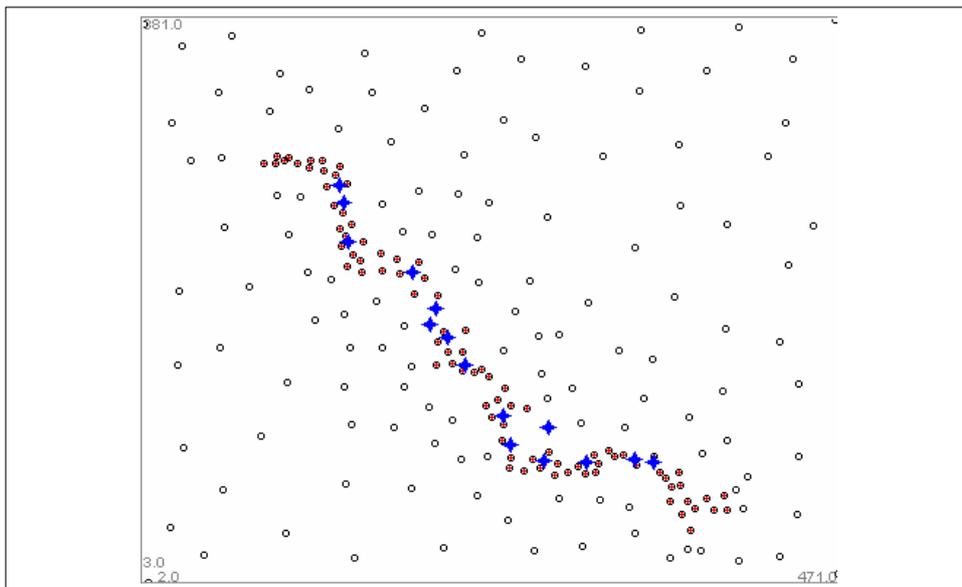


Рис. 17. Результат работы «Кристалла», $\gamma = 4,0$; $\beta = 1,2$

$$P_A(x) \geq \gamma P_X(x), \quad \forall x \in A.$$

Рост в соответствующей γ -версии глобального «Кристалла» описывается неравенствами

$$\text{Рост I: } P_{K_n}(x_{n+1}) \geq c(\gamma, n+1) P_{K_n}(x_{n+1}),$$

$$\text{Рост II: } \delta_{x_{n+1}}(x) \geq \frac{c(\gamma, n+1)(|X|-1)|K_n|}{|K_n|-1+c(\gamma, n+1)|K_n|} P_x(x) - (|K_n|-1)P_{K_n}(x),$$

где

$$c(\gamma, n+1) = \frac{\lambda |\overline{K}_{n+1}|}{|X| - 1 - |K_{n+1}| + \gamma}.$$

Вывод этих условий аналогичен доказательству утверждения 2.4.3.

8. На рис. 13–15 показана кристаллизация вокруг одной и той же основы соответственно при $\gamma = 1,0$; $\gamma = 1,2$ и $\gamma = 2,3$. На рис. 16, 17 приведены примеры работы «Кристалла» при $\gamma = 2,5$; $\beta = 1,2$ и $\gamma = 4,0$; $\beta = 1,2$.

ЛИТЕРАТУРА

1. *Гвишиани А.Д., Агаян С.М., Богоутдинов Ш.Р.* Математические методы геоинформатики. I. Новый подход к кластеризации // Кибернетика и системный анализ. — 2002. — № 2. — С. 104–122.
2. *Алгоритмы искусственного интеллекта для кластеризации магнитных аномалий / А.Д. Гвишиани, М. Диаман, В.О. Михайлов и др.* // Физика Земли. — 2002. — № 7. — С. 13–38.
3. *Application of artificial intelligence for Euler solution clustering / V.O. Mikhailov, A. Galdeano, M. Diamant et al.* // Geophysics. — 2003. — **68**, № 1. — P. 168–180.
4. *Gvishiani A., Dubois J.O.* Artificial intelligence and dynamic systems for geophysical applications. — Berlin: Springer, 2002. — 347 p.
5. *Математические методы геоинформатики. II. Алгоритмы нечеткой логики в задачах выделения аномалий на временных рядах / А.Д. Гвишиани, С.М. Агаян, Ш.Р. Богоутдинов и др.* // Кибернетика и системный анализ. — 2003. — № 4. — С. 103–111.
6. *Мандель И.Д.* Кластерный анализ. — М: Финансы и статистика, 1988. — 176 с.

Поступила 25.03.2004