УДК 004.032.26

Б.В. КОЛЧИГИН, Е.В. БОДЯНСКИЙ

АДАПТИВНАЯ НЕЧЕТКАЯ КЛАСТЕРИЗАЦИЯ С ПЕРЕМЕННЫМ ФАЗЗИФИКАТОРОМ

Ключевые слова: *нечеткая кластеризация, фаззификатор, нейронная сеть Кохонена, алгоритм самообучения.*

ВВЕДЕНИЕ

Задача кластеризации многомерных данных с целью нахождения в анализируемых массивах информации однородных в некотором смысле групп (классов, кластеров) наблюдений является важной составной частью разведочного анализа данных [1, 2]. Традиционно подход к ее решению предполагает, что каждое наблюдение относится только к одному кластеру, хотя более естественной представляется ситуация, когда обрабатываемый вектор признаков с различными уровнями принадлежности (вероятности, возможности) может относиться сразу к нескольким классам.

Данный вариант — предмет рассмотрения нечеткого (фаззи-) кластерного анализа [2–4], в основе которого лежит предположение, что классы однородных данных не разделяются, а перекрываются, и что каждому наблюдению может приписываться некоторый уровень принадлежности каждому кластеру в интервале от нуля до единицы [2].

Исходной информацией для рассматриваемой задачи является выборка наблюдений, сформированная из N n-мерных векторов признаков $x(1), x(2), \ldots$ $\ldots, x(k), \ldots, x(N)$, при этом для удобства численной реализации исходные данные предварительно центрируются и стандартизуются так, чтобы все наблюдения принадлежали гиперкубу $[-1, 1]^n$. Результатом кластеризации является разбиение исходного массива данных на m классов с некоторым уровнем принадлежности $u_j(k)$ k-го вектора признаков $x(k)$ j-му кластеру, $j = 1, 2, \ldots, m$.

В настоящей статье предлагаются адаптивные численно простые процедуры нечеткой кластеризации для рекуррентной обработки данных в on-line режиме по мере поступления наблюдений, предназначенные для работы в условиях априорной неопределенности относительно характера границ (нечеткости, размытости) между классами.

ВЕРОЯТНОСТНАЯ НЕЧЕТКАЯ КЛАСТЕРИЗАЦИЯ

В классе процедур нечеткой кластеризации наиболее строгими с математической точки зрения являются алгоритмы, основанные на целевых функциях [3] и решающие задачу их оптимизации при тех или иных априорных предположениях. Более распространен вероятностный подход, основанный на минимизации критерия (целевой функции)

$$E(u_j, c_j) = \sum_{k=1}^{N} \sum_{j=1}^{m} u_j^{\beta}(k) \| x(k) - c_j \|^2 \tag{1}$$

при ограничениях

$$\sum_{j=1}^{m} u_j(k) = 1, \tag{2}$$

$$0 \le \sum_{k=1}^{N} u_j(k) \le N, \tag{3}$$

где $u_j(k) \in [0,1]$ — уровень принадлежности вектора $x(k)$ j-му классу, c_j — центроид (прототип) j-го кластера, β — неотрицательный параметр фаззификации (фаззификатор), определяющий размытость границ между кластерами, $k = 1, 2, \ldots, N$. Результатом кластеризации является $(N \times m)$-матрица $U = \{u_j(k)\}$, называемая матрицей нечеткого разбиения.

Заметим, что в силу ограничения (2) элементы матрицы U рассматриваются как вероятности гипотез принадлежности векторов данных определенным кластерам, из-за чего процедуры, порождаемые минимизацией, называются вероятностными алгоритмами нечеткой кластеризации. При этом число кластеров m задается априорно и не может изменяться в процессе вычислений.

Вводя функцию Лагранжа

$$L(u_j(k), c_j, \lambda(k)) = \sum_{k=1}^{N} \sum_{j=1}^{m} u_j^{\beta}(k) \| x(k) - c_j \|^2 + \sum_{k=1}^{N} \lambda(k) \left(\sum_{j=1}^{m} u_j(k) - 1 \right) \tag{4}$$

(здесь $\lambda(k)$ — неопределенный множитель Лагранжа) и решая систему уравнений Каруша–Куна–Таккера, несложно получить искомое решение в виде

$$\begin{cases} u_j(k) = \dfrac{(\| x(k) - c_j \|^2)^{\frac{1}{1-\beta}}}{\sum_{l=1}^{m} (\| x(k) - c_l \|^2)^{\frac{1}{1-\beta}}}, \\[2ex] c_j = \dfrac{\sum_{k=1}^{N} u_j^{\beta}(k) x(k)}{\sum_{k=1}^{N} u_j^{\beta}(k)}, \\[2ex] \lambda(k) = -((\sum_{l=1}^{m} \beta \| x(k) - c_l \|^2)^{\frac{1}{1-\beta}})^{1-\beta}. \end{cases} \tag{5}$$

Оно совпадает при $\beta = 2$ с алгоритмом нечетких C-средних Бездека (Fuzzy C-Means, FCM) [3], а при $\beta \to 1$ близко к результатам, получаемым с помощью популярного алгоритма обычной (четкой) кластеризации K-средних (Hard K-Means, HKM) [5].

Алгоритм (5) нельзя в полном смысле назвать нечетким, поскольку он содержит четкое значение фаззификатора $1 < \beta < \infty$, выбираемое обычно из эмпирических соображений и существенным образом влияющее на получаемые результаты. При этом вариация β от единицы до бесконечности соответствует переходу от совершенно четких границ ($\beta \to 1$) до полного их размытия ($\beta \to \infty$), когда наблюдения с одинаковым уровнем принадлежности относятся ко всем кластерам [6]. Эффект влияния классификатора на «размытость» границы показан на рис. 1.

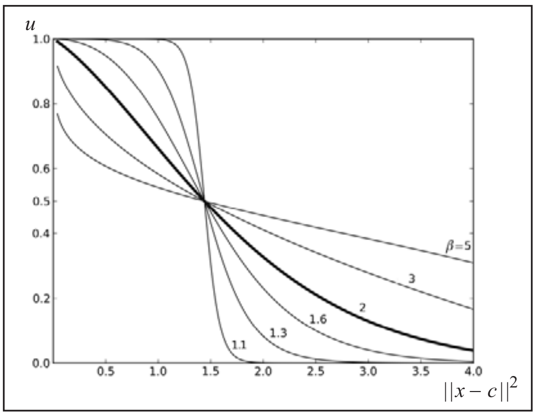


Рис. 1. Уровни принадлежности при разных значениях фаззификатора β

В связи с этим Клавонном и Хёппнером [7] для решения задачи нечеткой кластеризации вместо критерия (1) с четким фаззификатором предложено использовать выражение

$$E(u_j, c_j) = \sum_{k=1}^{N} \sum_{j=1}^{m} (\alpha u_j^2(k) + (1-\alpha) u_j(k)) \,||\, x(k) - c_j \,||^2 \tag{6}$$

с ограничениями (2), (3), где $0 < \alpha \le 1$ — настраиваемый параметр, определяющий характер получаемого решения.

Вводя функцию Лагранжа

$$L(u_j(k), c_j, \lambda(k)) = \sum_{k=1}^{N} \sum_{j=1}^{m} (\alpha u_j^2(k) + (1-\alpha)\, u_j(k)) \,||\, x(k) - c_j \,||^2 + \\ + \sum_{k=1}^{N} \lambda(k) \left(\sum_{j=1}^{m} u_j(k) - 1 \right) \tag{7}$$

и решая систему уравнений Каруша–Куна–Таккера

$$\begin{cases} \dfrac{\partial L(u_j(k), c_j, \lambda(k))}{\partial u_j(k)} = (2\alpha u_j(k) + 1 - \alpha) \,||\, x(k) - c_j \,||^2 + \lambda(k) = 0, \\ \nabla_{c_j} L(u_j(k), c_j, \lambda(k)) = -\sum_{k=1}^{N} 2\, (\alpha u_j^2(k) + (1-\alpha)\, u_j(k))\, (x(k) - c_j) = \vec{0}, \\ \dfrac{\partial L(u_j(k), c_j, \lambda(k))}{\partial \lambda(k)} = \sum_{j=1}^{m} u_j(k) - 1 = 0, \end{cases} \tag{8}$$

получаем систему

$$\begin{cases} u_j(k) = -\dfrac{1-\alpha}{2\alpha} + \dfrac{1 + m\dfrac{1-\alpha}{2\alpha}}{\sum_{l=1}^{m} \dfrac{||\, x(k) - c_j \,||^2}{||\, x(k) - c_l \,||^2}}, \\ c_j = \dfrac{\sum_{k=1}^{N} (\alpha u_j^2(k) + (1-\alpha)\, u_j(k)) x(k)}{\sum_{k=1}^{N} (\alpha u_j^2(k) + (1-\alpha) u_j(k))}, \\ \lambda(k) = -\dfrac{1 + m\dfrac{1-\alpha}{2\alpha}}{\sum_{l=1}^{m} (2\,\alpha \,||\, x(k) - c_l \,||^2\,)^{-1}}. \end{cases} \tag{9}$$

Несложно видеть, что при $\alpha = 1$ имеем FCM-алгоритм Бездека [3]

$$\begin{cases} u_j^{\mathrm{FCM}}(k) = \dfrac{||\, x(k) - c_j \,||^2}{\sum_{l=1}^{m} ||\, x(k) - c_l \,||^2}, \\ c_j^{\mathrm{FCM}} = \dfrac{\sum_{k=1}^{N} (u_j^{\beta}(k))^2\, x(k)}{\sum_{k=1}^{N} (u_j^{\beta}(k))^2}. \end{cases} \tag{10}$$

В результате первое соотношение (9) можно переписать в виде

$$u_j(k) = -\frac{1-\alpha}{2\alpha} + \left(1 + m\frac{1-\alpha}{2\alpha}\right) u_j^{\mathrm{FCM}}(k). \tag{11}$$

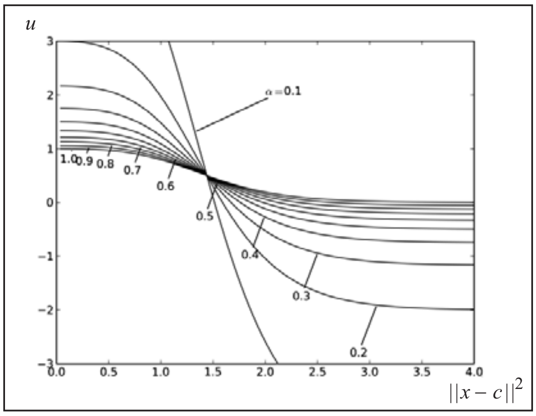


Рис. 2. Уровни принадлежности при разных значениях параметра α

На рис. 2 проиллюстрирован характер изменения уровней принадлежности в зависимости от значения параметра α, при этом отрицательные значения этих уровней не противоречат общим положениям теории систем нечеткого вывода и свидетельствуют о «непринадлежности» конкретного наблюдения данному кластеру [8, 9].

Заметим, что при $\alpha = \frac{1}{3}$ получаем компактное выражение

$$u_j(k) = (1+m) u_j^{\mathrm{FCM}}(k). \tag{12}$$

При использовании всех рассмотренных выше процедур подразумевается, что выборка, подлежащая кластеризации и содержащая N наблюдений, задана заранее и не может изменяться в процессе обработки. В [10-14] введена группа рекуррентных адаптивных процедур, позволяющих решать задачу кластеризации в on-line режиме.

Применяя к (7) процедуру нелинейного программирования Эрроу–Гурвица–Удзавы, получаем адаптивный алгоритм нечеткой кластеризации по критерию (6)

$$\begin{cases} u_j(k+1) = -\dfrac{1-\alpha}{2\alpha} + \dfrac{1 + m\dfrac{1-\alpha}{2\alpha}}{\displaystyle\sum_{l=1}^{m} \frac{\|x(k+1) - c_j(k)\|^2}{\|x(k+1) - c_l(k)\|^2}}, \\[2ex] c_j(k+1) = c_j(k) + \eta(k)\left(\alpha u_j^2(k+1) + (1-\alpha)\, u_j(k+1)\right)\left(x(k+1) - c_j(k)\right), \end{cases} \tag{13}$$

где $\eta(k)$ — параметр шага настройки. Несложно заметить, что второе соотношение (13) есть не что иное, как правило самообучения Кохонена [15] типа «Победитель получает больше» (WTM).

ВОЗМОЖНОСТНАЯ НЕЧЕТКАЯ КЛАСТЕРИЗАЦИЯ

Основные недостатки вероятностного подхода связаны с ограничением (2), требующим равенства единице суммы принадлежностей каждого наблюдения всем кластерам, и априорным заданием количества кластеров m. Этих недостатков в значительной мере лишены методы возможностной нечеткой кластеризации, изначально предложенные Кришнапурамом и Келлером [16].

В возможностных алгоритмах кластеризации целевая функция имеет вид

$$E(u_j, c_j) = \sum_{k=1}^{N}\sum_{j=1}^{m} u_j^{\beta}(k) \|x(k) - c_j\|^2 + \sum_{j=1}^{m} \mu_j \sum_{k=1}^{N} (1 - u_j(k))^{\beta}, \tag{14}$$

где скалярный параметр $\mu_j > 0$ определяет расстояние, на котором уровень принадлежности принимает значение 0,5, т.е. если

$$\|x(k) - c_j\|^2 = \mu_j, \tag{15}$$

то $u_j(k) = 0{,}5$.

Прямая минимизация (14) по $u_j(k)$ и c_j дает известное решение

$$\begin{cases} u_j(k) = \left(1 + \left(\dfrac{||x(k) - c_j||^2}{\mu_j}\right)^{\frac{1}{\beta - 1}}\right)^{-1}, \\ c_j = \dfrac{\sum\limits_{k=1}^{N} u_j^{\beta}(k) x(k)}{\sum\limits_{k=1}^{N} u_j^{\beta}(k)}, \\ \mu_j(k) = \dfrac{\sum\limits_{k=1}^{N} u_j^{\beta}(k) ||x(k) - c_j||^2}{\sum\limits_{k=1}^{N} u_j^{\beta}(k)}. \end{cases} \tag{16}$$

Использование вместо (14) критерия

$$E(u_j, c_j) = \sum_{k=1}^{N} \sum_{j=1}^{m} (\alpha u_j^2(k) + (1-\alpha) u_j(k)) ||x(k) - c_j||^2 + \sum_{j=1}^{m} \mu_j \sum_{k=1}^{N} (\alpha(1 - u_j(k)))^2 \tag{17}$$

ведет к возможностной процедуре с переменным фаззификатором:

$$\begin{cases} u_j(k) = \dfrac{(\alpha\mu_j + (1-\alpha)||x(k) - c_j||^2) + \mu_j}{2\alpha(||x(k) - c_j||^2 + \mu_j)}, \\ c_j = \dfrac{\sum\limits_{k=1}^{N} (\alpha u_j^2(k) + (1-\alpha) u_j(k)) x(k)}{\sum\limits_{k=1}^{N} (\alpha u_j^2(k) + (1-\alpha) u_j(k))}, \\ \mu_j = \dfrac{\sum\limits_{k=1}^{N} (\alpha u_j^2(k) + (1-\alpha) u_j(k)) ||x(k) - c_j||^2}{\sum\limits_{k=1}^{N} (\alpha u_j^2(k) + (1-\alpha) u_j(k))}. \end{cases} \tag{18}$$

В адаптивном варианте при обработке последовательно поступающих данных имеем рекуррентную процедуру

$$\begin{cases} u_j(k+1) = \dfrac{(\alpha\mu_j(k) + (1-\alpha)||x(k+1) - c_j(k)||^2) + \mu_j(k)}{2\alpha(||x(k+1) - c_j(k)||^2 + \mu_j(k))}, \\ c_j(k+1) = c_j(k) + \eta(k)(\alpha u_j^2(k+1) + (1-\alpha) u_j(k+1))(x(k+1) - c_j(k)), \\ \mu_j(k+1) = \dfrac{\sum\limits_{p=1}^{k+1} (\alpha u_j^2(p) + (1-\alpha) u_j(p)) ||x(p) - c_j(k+1)||^2}{\sum\limits_{p=1}^{k+1} (\alpha u_j^2(p) + (1-\alpha) u_j(p))}. \end{cases} \tag{19}$$

Заметим, что в (13) и (19) алгоритмы настройки центроидов кластеров совпадают, а отличие состоит в вычислениях уровней принадлежности к конкретному кластеру. Важно также, что возможностные процедуры, в отличие от вероятностных алгоритмов, позволяют в процессе обработки данных выявлять новые кластеры. Так, если уровень принадлежности наблюдения $x(k+1)$ ко всем кластерам окажется ниже некоторого наперед заданного порога, можно говорить о появлении $(m+1)$-го кластера с начальными координатами центроида $c_{m+1}=x(k+1)$.

НЕЧЕТКАЯ КЛАСТЕРИЗАЦИЯ С ПОДАВЛЕНИЕМ

В системах нечеткой кластеризации с подавлением [17] результаты, полученные с помощью нечетких процедур, модифицируются таким образом, что усиливается значимость результата с наивысшим уровнем принадлежности, остальные значения принадлежностей подавляются. Пусть для наблюдения $x(k)$ вычислены значения $u_1(k), u_2(k), \ldots, u_m(k)$, среди которых находится наибольшее значение $u_p(k)$. Далее это значение усиливается так, что

$$u_p^S(k)=1-\sigma\sum_{\substack{j=1\\ j\neq p}}^{m}u_j(k)=1-\sigma+\sigma u_p(k),\ \ 0\leq\sigma\leq 1, \tag{20}$$

остальные принадлежности подавляются в соответствии с выражением

$$u_j^S(k)=u_j(k)=\sigma u_j(k),\ j\neq p. \tag{21}$$

Если обобщить выражение (20) на возможностный случай, без ограничения (2) оно принимает вид

$$u_p^S(k)=\sum_{j=1}^{m}u_j(k)-\sigma\sum_{\substack{j=1\\ j\neq p}}^{m}u_j(k)=$$

$$=\sum_{j=1}^{m}u_j(k)-\sigma\left(\sum_{j=1}^{m}u_j(k)-u_p(k)\right)=\sigma u_p(k)+(1-\sigma)\sum_{j=1}^{m}u_j. \tag{22}$$

Несложно заметить, что с уменьшением параметра σ результаты становятся «более четкими», а с ростом σ они стремятся к решению, получаемому при фаззификаторе $\beta=2$. Видно, что параметр σ в (20), (21) имеет тот же смысл, что и α в целевой функции (6). В связи с этим целесообразно проводить в процессе обработки данных двойную коррекцию с различными параметрами α: первый раз на этапе собственно кластеризации, второй — на этапе подавления.

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

Алгоритм (19) реализован как алгоритм самообучения карты Кохонена. Эффективность полученной в результате адаптивной нейро-фаззи сети Кохонена для нечеткой кластеризация (Adaptive Fuzzy Kohonen Clustering Network, AFKCN) исследовалась на выборках из репозитория UCI [18] и сравнивалась с алгоритмом нечетких C-средних FCM в трех сериях эксперименов.

Для первой серии экспериментов использованы три размеченные выборки наблюдений. Выборка Iris включает 150 наблюдений, разделенных на три класса; каждое наблюдение содержит четыре признака. Выборка Wine включает 178 векторов наблюдений, разделенных на три класса; каждое наблюдение содержит 13 признаков. Выборка Wisconsin Diagnostic Breast Cancer (WDBC) включает 569 наблюдений, разделенных на два класса; каждое наблюдение содержит 30 признаков.

Поскольку для каждой выборки существуют метки правильной классификации, эффективность кластеризации измерялась в процентах точности относительно эталонного разбиения после дефаззификации. Сравнивались результаты работы метода четких K-средних (НКМ), нечетких C-средних при $\beta = 2$ (FCM) и AFKCN при разных значениях параметров. В табл. 1 приведены средний, максимальный и минимальный результаты для серий из 50 экспериментов на каждой выборке.

Таблица 1

Алгоритм	Результаты работы алгоритмов критеризации при выборках								
	Iris			Wine			WDBC		
	avg	max	min	avg	max	min	avg	max	min
НКМ	87	96	60	68	74	54	83	91	66
FCM ($\beta = 2$)	70	72	33	69	74	33	86	87	86
AFKCN без подавления	72	96	68	63	75	27	87	93	63
AFKCN с подавлением ($\alpha = \sigma = 0{,}1$)	85	93	66	70	73	66	89	91	83
AFKCN с подавлением ($\alpha = \sigma = 0{,}3$)	76	93	66	68	70	56	89	91	86
AFKCN с подавлением ($\alpha = \sigma = 0{,}5$)	75	90	66	68	70	66	90	91	88

Результаты показывают, что эффективность AFKCN с подавлением в общем выше и стабильнее, чем у алгоритмов НКМ и FCM. Параметр α оказывает тем большее влияние на эффективность, чем выше степень перекрытия классов, и почти незначим в случае линейно разделимых классов, как в выборке WDBC.

В следующей серии экспериментов исследовалась зависимость эффективности работы AFKCN от значения параметра α в режимах с подавлением и без подавления на искусственной выборке, состоящей из пяти частично пересекающихся классов по 200 нормально распределенных наблюдений в каждом. На графиках (рис. 3) показаны среднее значение точности в серии из 100 экспериментов для каждого значения α, а также средние положительные и отрицательные отклонения в каждой серии. Для режима с подавлением параметр подавления σ принят равным α. Для сравнения на третьем графике приведена зависимость точности кластеризации FCM от фаззификатора $1 < \beta < 5$.

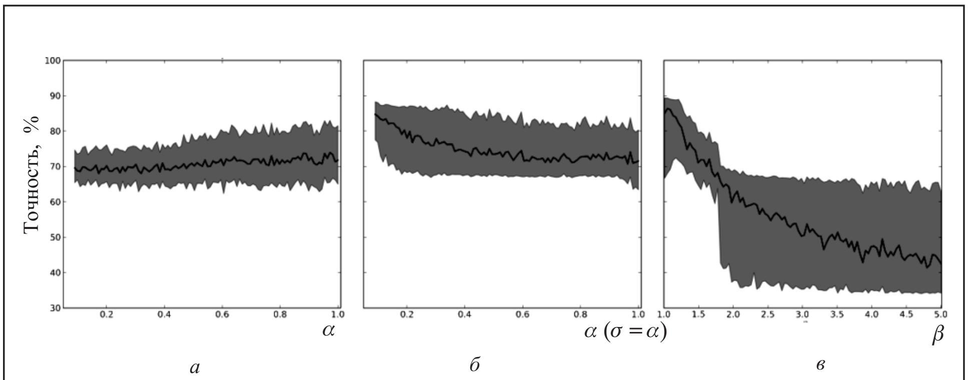


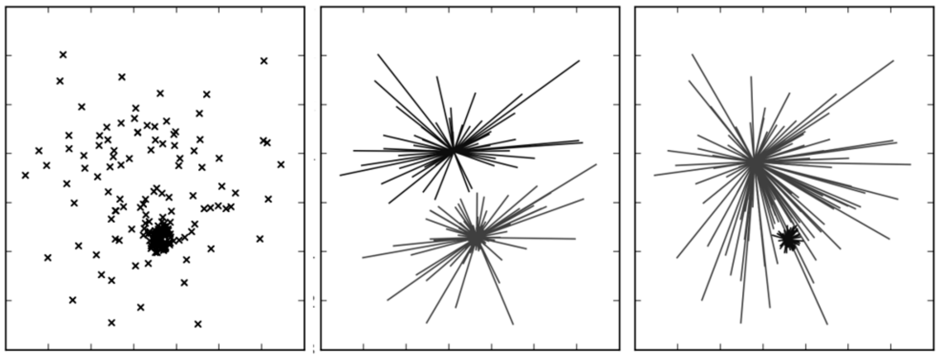
Рис. 3. Графики зависимости точности кластеризации от параметра α : AFKCN без подавления (*а*); AFKCN с подавлением (*б*); FCM (*в*)

В данном эксперименте подтвердилось, что для хорошо разделимой выборки значение параметра α не оказывает существенного влияния на итоговое разбиение, что выгодно отличает его от фаззификатора. С увеличением параметра α в режиме без подавления наблюдается увеличение дисперсии результатов при одном и том же значении α, однако на средний результат серии это не влияет.

При работе с подавлением характер влияния параметров AFKCN на результат подобен влиянию фаззификатора FCM, однако для AFKCN точность понижается значительно медленнее и остается в среднем выше, чем в режиме без подавления на всем диапазоне значений параметров α и σ.

В предыдущих экспериментах использованы равные параметры α и σ как на этапе кластеризации, так и этапе подавления. В следующем эксперименте продемонстрировано, что параметры α и σ целесообразно разделить, если в данных ожидаются кластеры существенно разной плотности. Этот результат может быть полезен при работе с сильно зашумленными данными, когда необходимо выделить выбросы или фон в отдельный кластер.

Для иллюстрации была создана синтетическая выборка, состоящая из 200 наблюдений, нормально распределенных относительно двух центров с разными параметрами распределения. AFKCN правильно проводит кластеризацию, если $\sigma < \alpha$. На рис. 4, *в* показан результат при $\alpha = 0,5, \sigma = 0,1$ (точность 97,5%), для сравнения на рис. 4, *б* приведен результат FCM при $\beta = 2$ (точность 71,5%).



а *б* *в*

Рис. 4. Результаты кластеризации на выборке с переменной плотностью наблюдений: выборка (*а*); FCM (*б*); AFKCN (*в*)

ЗАКЛЮЧЕНИЕ

В настоящей работе рассмотрена задача нечеткой кластеризации многомерных наблюдений с переменным фаззификатором и предложена группа адаптивных алгоритмов самообучения, позволяющих обрабатывать данные в реальном времени по мере их поступления. Введенные алгоритмы характеризуются численной простотой и обладают большой гибкостью при работе в условиях априорной неопределенности относительно характера распределения данных в кластерах. Дальнейшие исследования могут быть направлены на автоматическую настройку параметров нечеткости и подавления, а также настройку формы кластеров.

СПИСОК ЛИТЕРАТУРЫ

1. Тьюки Дж. Анализ результатов наблюдений. Разведочный анализ. — М.: Мир, 1981. — 693 с.
2. Höppner F., Klawonn F., Kruse R., Runkler T. Fuzzy clustering analysis: Methods for classification, data analysis and image recognition. — Chichester: John Willey & Sons, 1999. — 289 p.
3. Bezdek J.C. Pattern recognition with fuzzy objective function algorithms. — New York: Plenum Press, 1981. — 272 p.
4. Gath I., Geva A.B. Unsupervised optimal fuzzy clustering // Pattern Analysis and Machine Intelligence. — 1989. — **2**, N 7. — P. 773–787.
5. McQueen J. On convergence of *k*-means and partitions with minimum average variance // Ann. Math. Statist. — 1965. — **36**. — P. 1084.
6. Krishnapuram R., Keller J.M. The possibilistic C-means algorithm: Insights and recommendations // IEEE Trans. Fuzzy Systems. — 1996. — **4**, N 3. — P. 385–393.
7. Klawonn F., Höppner F. What is fuzzy about fuzzy clustering? Understanding and improving the concept of the fuzzifier // Lect. Notes Comp. Sci. — Berlin; Heidelberg: Springer, 2003. — **2811**. — P. 254–264.
8. Mitaim S., Kosko B. What is the best shape for a fuzzy set in function approximation? // Proc. 5th IEEE Int. Conf. on Fuzzy Systems “Fuzzy-96”. — New Orleans (USA), 1996. — **2**. — P. 1233–1237.
9. Mitaim S., Kosko B. Adaptive joint fuzzy sets for function approximation // Proc. Int. Conf. on Neural Networks “ICNN-97”. — Barcelona (Spain), 1997. — P. 537–542.
10. Park D. C., Dagher I. Gradient based fuzzy c-means (GBFCM) algorithm // Proc. IEEE Int. Conf. on Neural Networks. — San Diego (USA), 1984. — P. 1626–1631.
11. Chung F.L., Lee T. Fuzzy competitive learning // Neural Networks. — 1994. — **7**, N 3. — P. 539–552.
12. Бодянский Е.В., Горшков Е.В., Кокшенев И.В., Колодяжный В.В. Об адаптивном алгоритме нечеткой кластеризации данных // Адаптивні системи автоматичного управління. — Дніпропетровськ: Системні технології, 2002. — Вип. 5(25). — С. 108–117.
13. Bodyanskiy Ye., Kolodyaznhiy V., Stephan A. Recursive fuzzy clustering algorithms // Proc. 10th East West Fuzzy Colloqium. — Zittau (Germany), 2002. — P. 276–283.
14. Bodyanskiy Ye. Computational intelligence techniques for data analysis // Lect. Notes in Informatics. — Bonn (Germany), 2005. — **P-72**. — P. 15–36.
15. Kohonen T. Self-organizing maps. — Berlin: Springer-Verlag, 1995. — 362 p.
16. Krishnapuram R., Keller J. M. A possibilistic approach to clustering // Fuzzy Systems. — 1993. — **1**, N 2. — P. 98–110.
17. Fan J.-L., Zhen W.-Z., Xie W.-X. Supressed fuzzy C-means clustering algorithm // Pattern Recognition Letters. — 2003. — **23**. — P. 1607–1612.
18. Frank A., Asuncion A. UCI Machine learning repository. — Irvine (CA): Univ. of California, School of Information and Computer Science, 2010. — http://archive.ics.uci.edu/ml.

Поступила 06.08.2012