

## КЛАСТЕРНЫЙ АНАЛИЗ ДЛЯ МНОГОМЕРНЫХ ОБЪЕКТОВ В УСЛОВИЯХ НЕЧЕТКИХ ДАННЫХ

Ю.А. ЗАК

**Аннотация.** Приведены описанные в литературе множества различных областей практических приложений многофакторного кластерного анализа в условиях нечетких исходных данных. Предложены новые алгоритмы и формульные выражения объединения различных многомерных объектов, параметры которых заданы fuzzy-множествами, в кластеры и вычисления координат центроидов их функций принадлежности. Сформулированы различные виды критериев кластеризации в виде минимизации средневзвешенной и представленной в различной метрике суммы расстояний между центроидами объектов и кластеров, а также максимизации расстояний между центроидами различных кластеров. Предложены постановки и математические модели трех различных *NP*-сложных задач многомерной кластеризации в условиях fuzzy-данных, при решении которых может быть использован любой из рассмотренных критериев оптимальности. Разработаны эвристические алгоритмы приближенного решения двух сформулированных задач. Алгоритм решения первой задачи проиллюстрирован на числовом примере. Полученные результаты могут послужить направлением дальнейших исследований и найти широкое практическое применение.

**Ключевые слова:** кластерный анализ, многомерные функции принадлежности, центроиды fuzzy-множеств объектов и кластеров, центры тяжести и середины сечений нечетких множеств, критерии оптимальности и алгоритмы кластеризации.

### ВВЕДЕНИЕ

Кластерный анализ — это разбиение заданного множества, состоящего из  $N$  многомерных объектов ( $i = 1, \dots, N$ ), на некоторое количество подмножеств (групп), называемых кластерами. При этом каждый кластер должен включать близкие по свойствам, характеристикам или своим значениям объекты, а объекты разных кластеров существенно отличались друг от друга [1–4; 12]. Количество таких кластеров  $O_k$  равно  $K$  ( $k = 1, \dots, K$ ), где  $K \ll N$ . Количество кластеров может быть задано или определяться в процессе реализации алгоритма при возможных заданных граничных значениях на количества кластеров. Каждый из объектов может определяться одним параметром, представленным числовым или качественным значением, либо некоторым вектором параметров в  $m$ -мерном пространстве  $x_i^j$ ,  $j = 1, \dots, m$ ,  $i = 1, \dots, N$ . Тогда каждый кластер также представлен некоторым множеством точек в этом  $m$ -мерном векторном пространстве. Так как разные параметры (свойства) объектов могут быть представлены в различных единицах и различных диапазонах изменения, что может исказить или даже дать неправильный результат, необходимо привести эти данные к одним и тем же единицам измерения. Это достигается применением коэффициентов масштабирования

значений отдельных свойств. Степень важности различных свойств учитывается введением соответствующих весовых коэффициентов.

Выборка таких объектов должна быть однородной, т.е. не содержать не характерных для данного класса задач «выбросов».

Степень сходства между данными из одного кластера должна быть как можно более высокой, а между данными из разных кластеров — как можно меньшей, в связи с чем концепция подобия имеет здесь решающее значение. В качестве степени сходства между данными из одного кластера могут быть использованы евклидово расстояние, а также суммарное квадратичное отклонение объектов кластера от их центров (центроидов). Квадратичное евклидово расстояние способствует увеличению контрастности кластеров. Расстояния между центроидами данных различных кластеров, которые должны быть максимизированы, могут также определяться суммарными средне-взвешенными квадратичными или линейными значениями. Каждый объект  $A_i$  и каждый кластер  $O_k$  являются нечеткими множествами. Алгоритм кластеризации — это определение функции или некоторого решающего правила, на основе которого каждому объекту ставится в соответствие номер кластера, к которому этот объект принадлежит.

Область применений кластерного анализа очень широка. Эти методы используются в медицине, технической диагностике, психологии, биологии, археологии, маркетинге и во многих других дисциплинах [9–19].

В медицине пациенты со схожими характеристиками и симптомами по результатам иерархического кластерного анализа могут быть сгруппированы таким образом, чтобы можно было определить наличие и степень некоторого заболевания и начать соответствующие лечебные меры в соответствии с установленным диагнозом (см., например, [10, 16, 17, 19]). Примеры использования кластерного анализа в экономике для классификации ситуаций приведены в работе [9]. В системе государственного управления кластерный анализ позволяет распределить налогоплательщиков по категориям внимания, к которым будут применяться соответствующие комплексы мероприятий [13]. В маркетинге можно определить сегменты клиентов с аналогичными моделями покупательского поведения, а также перечни продуктов, пользующихся наибольшим спросом среди каждой из групп покупателей. Это даст возможность координации и повышения эффективности маркетинговой деятельности конкретно с определенными сегментами покупателей, что, в свою очередь, приведет к увеличению объема продаж. В проблемах защиты окружающей среды по результатам мониторинга могут быть определены зоны в воздухе с определенными степенями загрязнения, что даст возможность принять любые необходимые мероприятия по улучшению ситуации. В биологии с помощью кластерного анализа анализируются сложные сети взаимодействующих генов, состоящие порой из сотен или даже тысяч элементов. В информатике кластерный анализ используется для «интеллектуальной» группировки результатов, распознавания изображений [11] при поиске файлов и веб-сайтов, а также других информационных объектов, предоставляя пользователю возможность быстрой навигации для ускорения поиска необходимой ему информации.

Большое количество практических приложений основаны на том, что решения принимаются, опираясь на нечеткие данные. Разные параметры

или свойства объектов могут быть представлены нечеткими подмножествами с функциями принадлежности вида (см., например, [7, 8, 14, 15]):

$$\{x, [\mu_x(x)]\}; \mu_{Y_i(C_k)}[Y_i(C_k)] = \sum_{x \in C_k} \gamma_i [\mu_{x_i}(x_i)],$$

где  $\gamma_i$  — весовые коэффициенты, которые в частном случае могут быть равны 1.

Если при так называемой четкой или жесткой кластеризации каждый объект может принадлежать только одному кластеру, то при нечеткой кластеризации каждый объект с различными значениями функции принадлежности может одновременно принадлежать сразу нескольким кластерам. Одним из наиболее широко используемых алгоритмов нечеткой кластеризации является алгоритм (FCM) [1, 7, 12]). Любой объект  $x_i$  имеет некоторый набор данных, определяющих его принадлежность к некоторому кластеру. В алгоритме нечетких  $C$ -средних центроид кластера  $O_k - C_k$  является средним значением всех точек в многомерном пространстве, взвешенных по степени их принадлежности к данному кластеру:

$$C_k = \frac{\sum_{x_i \in C_k} [\mu_x(x), x]^r x}{\sum_{x_i \in C_k} [\mu_x, x]^r}.$$

Здесь  $r$  — показатель степени, с увеличением которого растет показатель нечеткости кластера.

Алгоритм FCM минимизирует функционал вида

$$F = \min \sum_{k=1}^K \sum_{i=1}^{C_k} [\mu_x(x), x]^r (x_i - z_k)^2.$$

Этот алгоритм осуществляет распределение  $N$  объектов, заданных нечеткими множествами, по  $K$  определяемым нечеткими множествами кластерам в соответствии с выбранным критерием оптимальности. Эффективность получения результата повышается при реализации алгоритма не с заранее заданным, а с автоматическим определяемым количеством кластеров.

## ПАРАМЕТРЫ КЛАСТЕРОВ, ВКЛЮЧАЮЩИЕ ОБЪЕКТЫ, ПРЕДСТАВЛЕННЫЕ НЕЧЕТКИМИ МНОЖЕСТВАМИ

Рассматриваются объекты, параметры или свойства которых, а также параметры кластеров представлены нечеткими подмножествами с функциями принадлежности вида

$$A_i = \{x_i, \mu_{A_i}(x_i) \mid x_i \in X_i\}, \mu_{A_i}(x_i) \in [0,1],$$

где

$$\int_{\min X_i}^{\max X_i} \mu_{A_i}(x_i) = 1, i = 1, \dots, N; \quad (1)$$

$$O_k = \{z_k, \mu_{O_k}(z_k) \mid z_k \in Z_k\},$$

где 
$$\int_{\min Z_k}^{\max Z_k} \mu_{Z_k}(z_k) = 1, \quad k = 1, \dots, K. \quad (2)$$

Здесь  $i = 1, \dots, N$  — индексы,  $N$  — количество рассматриваемых объектов;  $k = 1, \dots, K$ , — индексы и количество априори рассматриваемых кластеров.

Пусть нечеткое множество кластера  $O_k$ , включающего некоторое подмножество объектов  $A_i \in O_k$ , определяется согласно выражению

$$\mu_{O_k}(z_k) = \max_{A_i \in O_k} \{\mu_{A_i}(x_i) | x \in X_i\}, \quad k = 1, \dots, K.$$

Образование функции принадлежности такого кластера (ABCDEFHG), включающего два объекта иллюстрирует рис. 1.

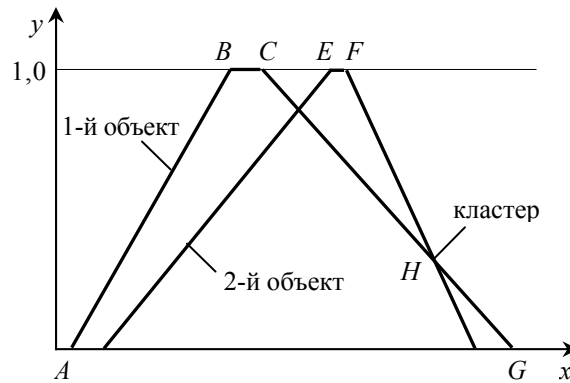


Рис. 1. Функция принадлежности нечетких множеств двух объектов и кластера

Каждый объект может определяться некоторым множеством свойств (параметров)  $x_i^j$ , количество которых равно  $m$ , значение каждого из которых определяется функцией принадлежности  $\mu_{A_i^j}(x_i^j)$ ,  $x_i^j \in X_i^j$ . Кластеры в этом случае также являются многомерными и выражены вектором функций принадлежности той же размерности  $m$ . Обобщенная функция принадлежности этого объекта и кластера  $O_k$  могут быть вычислены, например, по формулам:

$$\{x_i, \mu_{A_i}(x_i) | x_i \in X_i\} = \sum_{j=1}^m \alpha^j \{x_i^j, \mu_{A_i^j}(x_i^j) | x_i^j \in X_i^j\}, \quad i = 1, \dots, N; \quad (3)$$

$$\{z_k, \mu_{Z_k}(z_k) | z_k \in Z_k\} = \sum_{j=1}^m \alpha^j \{z_k^j, \mu_{Z_k^j}(z_k^j) | z_k^j \in Z_k^j\}, \quad k = 1, \dots, K. \quad (4)$$

Здесь  $\alpha^j$  — весовые коэффициенты, определяющие значимость каждого из этих свойств, величины которых удовлетворяют соотношениям

$$0 \leq \alpha^j \leq 1, \quad j = 1, \dots, m, \quad \sum_{j=1}^m \alpha^j = 1.$$

В качестве основной характеристики как отдельных объектов, так и каждого из кластеров выберем координату абсцисс центра тяжести соответствующего fuzzy-множества, которые рассчитываются по формулам (см., например, [5, 6]):

$$g_i^j = \int_{\min x_i^j}^{\max x_i^j} \frac{x_i^j \mu_{A_i^j}(x_i^j)}{\mu_{A_i^j}(x_i^j)}, \quad i=1, \dots, m; \quad g_i = \int_{\min X_i}^{\max X_i} \frac{x_i \mu_{A_i}(x_i)}{\mu_{A_i}(x_i)}, \quad i=1, \dots, N; \quad (5)$$

$$G_k^j = \int_{\min Z_k^j}^{\max Z_k^j} \frac{z_k^j \mu_{Z_k^j}(z_k^j)}{\mu_{Z_k^j}(z_k^j)}, \quad j=1, \dots, m, \quad G_k = \int_{\min Z_k}^{\max Z_k} \frac{z_k \mu_{Z_k}(z_k)}{\mu_{Z_k}(z_k)}, \quad k=1, \dots, K. \quad (6)$$

Ограничимся рассмотрением нечетких множеств объектов с функцией принадлежности треугольного или трапециевидального вида (рис. 2):

$$\mu_{A_i}(x_i) = (a_i, c_i, b_i), \quad \mu_{A_i}(x_i) = (a_i, c_{1i}, c_{2i}, b_i),$$

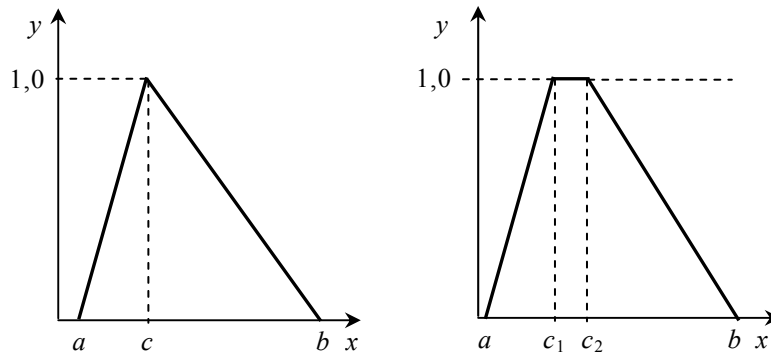


Рис. 2. Функции принадлежности fuzzy-множеств треугольного и трапециевидного типа

$$\mu_{A_k}(z_k) = (a_k, c_k, b_k), \quad \mu_{A_k}(z_k) = (a_k, c_{1k}, c_{2k}, b_k),$$

$$Z_k^{p+1} = [a(Z_k^{p+1}), c_1(Z_k^{p+1}), c_2(Z_k^{p+1}), b(Z_k^{p+1})],$$

$$a(Z_k^{p+1}) = \min[a_i, a(Z_k^p)], \quad b(Z_k^{p+1}) = \max[b_i, b(Z_k^p)], \quad (7)$$

$$c_1(Z_k^{p+1}) = \min[c_{1i}, c_{1i}(Z_k^p)], \quad c_2(Z_k^{p+1}) = \max[c_{2i}, c_{2i}(Z_k^p)], \quad (8)$$

$$c_1(Z_k^{p+1}) = \min[c_i, c_{1i}(Z_k^p)], \quad c_2(Z_k^{p+1}) = \max[c_i, c_{2i}(Z_k^p)]. \quad (9)$$

Формулы для вычисления значения функции принадлежности для этих fuzzy-множеств приведены, например, в работе [5]. Координаты оси абсцисс центра тяжести для треугольных fuzzy-множеств могут вычислены по простым формулам [5, 6]:

$$G_1(A_i) = \frac{1}{3}[a(A_i) + c(A_i) + b(A_i)], \quad G_2(A_k) = \frac{1}{3}[a(A_k) + c(A_k) + b(A_k)]. \quad (10)$$

Для трапециевидных fuzzy-множеств в качестве координаты  $G_1(O_k)$  может быть принято значение, вычисленное по следующему алгоритму.

Вычислим площади соответствующих частей fuzzy-множества: двух треугольников, которые обозначим  $S_1(O_k)$  и  $S_3(O_k)$  и прямоугольника, которую обозначим как  $S_2(O_k)$ , а также координаты оси абсцисс центров

тяжести этих частей, которые соответственно равны  $G[S_1(O_k)]$ ,  $G[S_3(O_k)]$  и  $G[S_2(O_k)]$ .

Определим весовые коэффициенты

$$\gamma_1 = \frac{S_1(O_k)}{S_1(O_k) + S_2(O_k) + S_3(O_k)}; \quad \gamma_2 = \frac{S_2(O_k)}{S_1(O_k) + S_2(O_k) + S_3(O_k)};$$

$$\gamma_3 = \frac{S_3(O_k)}{S_1(O_k) + S_2(O_k) + S_3(O_k)},$$

где  $\gamma_1 + \gamma_2 + \gamma_3 = 1$ .

Тогда координаты оси абсцисс центра тяжести трапеции равна

$$G(O_k) = \gamma_1 G[S_1(O_k)] + \gamma_2 G[S_2(O_k)] + \gamma_3 G[S_3(O_k)]. \quad (11)$$

Если объект  $A_i$  с треугольной функцией принадлежности включается в кластер  $O_k$  с трапециевидальной функцией принадлежности, то параметры функции принадлежности кластера изменяются в соответствии со следующими выражениями:

$$a(O_k) = \begin{cases} a(O_k), & \text{if } a(O_k) \leq a(A_i), \\ a(A_i), & \text{if } a(O_k) > a(A_i); \end{cases} \quad b(O_k) = \begin{cases} b(O_k), & \text{if } b(O_k) \geq b(A_i), \\ b(A_i), & \text{if } b(O_k) < b(A_i); \end{cases}$$

$$c_1(O_k) = \begin{cases} c_1(O_k), & \text{if } c_1(O_k) \leq c(A_i), \\ a(A_i), & \text{if } c_1(O_k) > c(A_i); \end{cases} \quad c_2(O_k) = \begin{cases} c_2(O_k), & \text{if } c_2(O_k) \geq c(A_i), \\ c(A_i), & \text{if } c_2(O_k) < c(A_i). \end{cases}$$

Если первые два объекта  $A_i$  и  $A_v$  с треугольными функциями принадлежности объединяются в один кластер, то параметры этого кластера определяются по формулам:

$$a(O_k) = \begin{cases} a(A_v), & \text{if } a(A_v) \leq a(A_i), \\ a(A_i), & \text{if } a(A_v) > a(A_i); \end{cases} \quad b(O_k) = \begin{cases} b(A_v), & \text{if } b(A_v) \geq b(A_i), \\ b(A_i), & \text{if } b(A_v) < b(A_i); \end{cases}$$

$$c_1(O_k) = \begin{cases} c(A_v), & \text{if } c(A_v) \leq c(A_i), \\ a(A_i), & \text{if } c(A_v) > c(A_i); \end{cases} \quad c_2(O_k) = \begin{cases} c(A_v), & \text{if } c(A_v) \geq c(A_i), \\ c(A_i), & \text{if } c(A_v) < c(A_i). \end{cases}$$

В качестве другой характеристики каждого fuzzy-множества в задачах кластерного анализа может быть принята средневзвешенная координат оси абсцисс средин сечений этих множеств при различных значениях ординаты функции принадлежности (рис. 3), т.е. точек на прямой  $Z_1 - Z_2$ .

Обозначим координаты оси абсцисс середины соответствующих сечений  $x(\mu_\tau)$ . Здесь  $\mu_\tau$  — значение координаты  $y$  в этом сечении. Как правило,

$$0 \leq \mu_{\tau_0} < \mu_{\tau_1} < \mu_{\tau_2} < \dots < \mu_{\tau_q} < \dots < \mu_{\tau_Q} \leq 1,0.$$

Тогда 
$$x(Z) = \sum_{q=0}^Q \delta_q x(\mu_q), \quad (12)$$

где  $0 \leq \delta_q \leq 1,0$  — весовые коэффициенты, причем  $\sum_{q=0}^Q \delta_q = 1,0$ .

В частном случае  $\delta_q = 1, 0; q = 0, 1, \dots, Q$ .

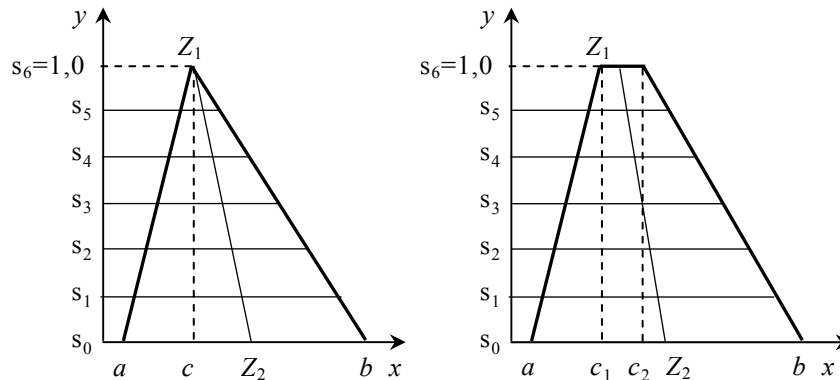


Рис. 3. Горизонтальные сечения функций принадлежности fuzzy-множеств

### КРИТЕРИИ ОПТИМАЛЬНОСТИ

В качестве критериев оптимальности при отнесении множества, включающего  $N$  объектов к одному из  $K$  кластеров, могут быть приняты параметры расстояний в различных метриках между центрами тяжести объектов и кластеров, а также между центроидами кластеров  $G_k$  и  $G_l$ .

$$F_1 = \sum_{k=1}^K \frac{1}{N_k} \sum_{A_i \in O_k} \tau_{ki}^1 |G_k - g_i| \rightarrow \min; \quad (13)$$

$$F_2 = \sum_{k=1}^K \frac{1}{N_k} \sum_{A_i \in O_k} \tau_{ki}^2 (G_k - g_i)^2 \rightarrow \min; \quad (14)$$

$$F_3 = \sum_{k=1}^K \max_{A_i \in O_k} \tau_{ki}^3 |G_k - g_i| \rightarrow \min; \quad (15)$$

$$F_4 = \sum_{k=1}^K \sum_{l=k+1}^K \pi_{kl}^1 |G_k - G_l| \rightarrow \max; \quad (16)$$

$$F_5 = \sum_{k=1}^K \sum_{l=k+1}^K \pi_{kl}^2 (G_k - G_l)^2 \rightarrow \max; \quad (17)$$

$$F_6 = \sum_{k=1}^K \sum_{l=k+1}^K \max_{(k,l)} \pi_{kl}^3 |G_k - G_l| \rightarrow \max. \quad (18)$$

В выражениях (13)–(18) и в дальнейшем  $G_k$  и  $g_i$  — соответственно координаты центров тяжести оси абсцисс центроидов  $k$ -го кластера и  $i$ -го объекта, а также между центроидами кластеров  $G_k$  и  $G_l$ . Здесь  $\tau_{ki}^p$ ,  $0 < \tau_{ki}^p \leq 1$   $p = 1, 2, 3$  — весовые коэффициенты, сумма которых в каждом из этих выражений равна  $\sum_{A_i \in O_k} \tau_{ki}^p = 1$ ,  $k = 1, \dots, K$ . Весовые коэффициенты

$0 < \pi_{kl}^p \leq 1$ ,  $p = 1, 2, 3$ ,  $k = 1, \dots, K-1$ ,  $l = 2, \dots, K$ ,  $\sum_{l=k+1}^K \pi_{kl}^p = 1$ ;  $N_k$  — количество объектов, включенных в кластер  $O_k$ .

Кроме того, могут использоваться многокритериальные подходы принятия решений с использованием минимизации аддитивной свертки критериев, как, например,

$$\Phi_1 = \alpha_1 \frac{1}{K} F_1 - \alpha_2 \frac{1}{K(K-1)} F_4; \quad \Phi_2 = \alpha_3 \frac{1}{K} F_2 - \alpha_4 \frac{1}{K(K-1)} F_5; \quad (19)$$

$$\Phi_3 = \alpha_5 \frac{1}{K} F_3 - \alpha_6 \frac{1}{K(K-1)} F_6, \quad (20)$$

а также

$$\Phi_4 = \alpha_7 \frac{1}{K} F_1 - \alpha_8 \frac{1}{K(K-1)} F_6; \quad \Phi_5 = \alpha_9 \frac{1}{K} F_3 - \alpha_{10} \frac{1}{K(K-1)} F_4; \quad (21)$$

$$\Phi_6 = \alpha_{11} \frac{1}{K} F_2 - \alpha_{12} \frac{1}{K(K-1)} F_4; \quad \Phi_7 = \alpha_{12} \frac{1}{K} F_3 - \alpha_{13} \frac{1}{K(K-1)} F_5; \quad (22)$$

$$\Phi_8 = \alpha_{14} \frac{1}{K} F_2 - \alpha_{15} \frac{1}{K(K-1)} F_6; \quad \Phi_9 = \alpha_{16} \frac{1}{K} F_1 - \alpha_{17} \frac{1}{K(K-1)} F_6. \quad (23)$$

Здесь  $\alpha_p$ ,  $0 < \alpha_p \leq 1$ .  $p = 1, 2, \dots, 17$  — весовые коэффициенты, сумма которых в каждом из этих выражений равна 1.

Если каждый объект представлен некоторым вектором параметров в  $m$ -мерном пространстве  $x_i^j$ ,  $j = 1, \dots, m$ ,  $i = 1, \dots, N$ , то в критерии оптимальности кластеризации могут быть представлены в виде:

$$E_1 = \min \frac{1}{N_k} \sum_{X_i \in O_k} \sum_{j=1}^m \beta_1^j |z_k^j - x_i^j|, \quad E_2 = \min \frac{1}{N_k} \sum_{X_i \in O_k} \sum_{j=1}^m \beta_2^j (z_k^j - x_i^j)^2,$$

$$E_3 = \min \frac{1}{N_k} \sum_{X_i \in O_k} \max_{1 \leq j \leq m} \beta_3^j |z_k^j - x_i^j|; \quad (23)$$

$$\Phi_1 = \max \left\{ \frac{1}{K(K-1)} \sum_{k=1}^K \sum_{l=k+1}^K \sum_{j=1}^m \delta_1^j |z_k^j - z_l^j| \right\},$$

$$\Phi_2 = \max \left\{ \frac{1}{K(K-1)} \sum_{k=1}^K \sum_{l=k+1}^K \sum_{j=1}^m \delta_2^j (z_k^j - z_l^j)^2 \right\},$$

$$\Phi_3 = \max \left\{ \frac{1}{K} \sum_{k=1}^K \sum_{l=k+1}^K \sum_{j=1}^m \max_{1 \leq j \leq m} \delta_3^j |z_k^j - z_l^j| \right\}. \quad (24)$$

Здесь  $\beta_p^j$ ,  $0 < \beta_p^j \leq 1$ .  $p = 1, 2, 3$ ; и  $\delta_r^j$ ,  $0 < \delta_r^j \leq 1$ .  $r = 1, 2, 3$ ; — весовые коэффициенты, сумма которых в каждом из этих выражений равна 1.

Отметим, что в выше приведенных выражениях, так и в других критериях и алгоритмах принятия решений задач кластеризации в условиях нечетких данных, вместо значений координаты абсцисс центра тяжести нечетких множеств могут использоваться так и середины соответствующих сечений этих fuzzy-множеств.



## ПОСТАНОВКИ ЗАДАЧ КЛАСТЕРИЗАЦИИ

Приведенные ниже постановки задач предусматривают работу с многомерными объектами и кластерами, параметры которых представлены нечеткими множествами. Рассмотрим три различные постановки задачи в условиях нечетких данных.

**Задача 1.** Определены функции принадлежности всех характеристик (признаков) каждого из  $K$  кластеров,  $O_k = \{z_k, \mu_{O_k}(z_k) | z_k \in Z_k\}$ ,  $\mu_{O_k}(z_k) \in [0, 1]$ ,  $k=1, \dots, K$ , а также известны функции принадлежности всех характеристик  $N$  многомерных объектов  $A_i = \{x_i, \mu_{A_i}(x_i) | x_i \in X_i\}$ ,  $\mu_{A_i}(x_i) \in [0, 1]$ ,  $i=1, \dots, N$ . Необходимо в соответствии с одним из выше приведенных критериев оптимальности распределить это множество объектов среди этих  $K$  кластеров.

Такая постановка задача очень характерна для медицинской диагностики, когда определены различные факторы отсутствия наличия и степени серьезности различных диагнозов и заболеваний, для технической диагностики, когда установлены признаки различного вида неисправностей, а также в задачах определения и классификации аварийных ситуаций причин и последствий экологических катастроф.

**Задача 2.** Известны функции принадлежности всех характеристик (признаков)  $N$  многомерных объектов  $A_i = \{x_i, \mu_{A_i}(x_i) | x_i \in X_i\}$ ,  $\mu_{A_i}(x_i) \in [0, 1]$ ,  $i=1, \dots, N$ . Установлено только количество кластеров  $K$ . Необходимо в процессе решения задачи определить функции принадлежности всех характеристик (признаков) каждого из  $K$  кластеров,  $O_k = \{z_k, \mu_{O_k}(z_k) | z_k \in Z_k\}$ ,  $\mu_{O_k}(z_k) \in [0, 1]$ ,  $k=1, \dots, K$ , и отнести каждый из объектов  $A_i$ ,  $i=1, \dots, N$ , только к одному из этих кластеров в соответствии с одним из выше приведенных критериев оптимальности (19)–(24).

**Задача 3.** Известны только функции принадлежности всех характеристик (признаков)  $N$  многомерных объектов  $A_i$ ,  $i=1, \dots, N$ . Количество выделяемых кластеров заранее не определено и может быть только ограничено некоторым числом  $B$ . Необходимо определить количество кластеров  $1 \leq K \leq B$ , а также функции принадлежности всех характеристик (признаков) каждого из этих  $K$ , отнести каждый из объектов  $A_i$ ,  $i=1, \dots, N$ , только к одному из этих кластеров в соответствии с одним из выше приведенных комплексных критериев оптимальности (19)–(24).

Отметим, что задачи 2 и 3 являются задачами  $NP$ -сложности и для решения этих задач в условиях большой размерности могут быть предложены только эвристические алгоритмы приближенного решения

## АЛГОРИТМ КЛАСТЕРНОГО АНАЛИЗА

Рассмотрим алгоритмы решения двух различных выше сформулированных задач. В двух описанных алгоритмах А.1 и А.2 решения задачи 1 используется один из критериев оптимальности  $F_1 - F_6$ , приведенных в формулах (13)–(18).

В начале работы алгоритмов приведем единицы и диапазоны изменения значений всех параметров объектов и кластеров (координат оси абсцисс этих параметров) к одним и тем же безразмерным единицам измерения:

$$X_{\min}^j = \min[\min_{1 \leq i \leq N} x_i^j; \min_{1 \leq k \leq K} x_k^j], \quad X_{\max}^j = \max[\max_{1 \leq i \leq N} x_i^j; \max_{1 \leq k \leq K} x_k^j],$$

$$X^j = (X_{\max}^j - X_{\min}^j).$$

Обозначим:  $\bar{X}_-^j = X_{\min}^j$ ,  $\bar{X}_+^j = X_{\max}^j$ . Тогда  $\bar{x}_i^j = \bar{X}_-^j + \frac{x_i^j}{X^j}$ ,  $i = 1, \dots, N$ ;

$$\bar{x}_k^j = \bar{X}_-^j + \frac{x_k^j}{X^j}, \quad k = 1, \dots, K.$$

### Алгоритм А.1.

1. Вычислим координаты абсцисс центров тяжести или средневзвешенную координату середин оси абсцисс выбранных сечений функций принадлежности нечетких множеств для каждого из признаков всех объектов и центров кластеров в соответствии с формулами (5)–(10) или (12).

2. Вычислим значения:

$$w_{ki}^j = |G_k^j - g_i^j|, \quad j = 1, \dots, m, \quad i = 1, \dots, N, \quad k = 1, \dots, K;$$

$$\bar{W}_{ki} = \sum_{j=1}^m \rho^j w_{ki}^j, \quad i = 1, \dots, N, \quad k = 1, \dots, K.$$

Объект  $A_i$  относится к кластеру  $O_{\bar{k}}$ , для которого значение  $\bar{W}_{ki}$  минимально, т.е.  $\bar{k}(A_i) = \arg \min_{1 \leq k \leq K} \bar{W}_{ki}$ .

Другой алгоритм решения задачи 1 отличается от описанного выше тем, что в процессе решения производится пересчет координаты абсцисс центроидов вновь преобразованных кластеров.

### Алгоритм А.2.

1. Выполняем пункты 1 и 2 алгоритма А.1.

2. Дальнейшие вычисления предусматривают выполнение  $N$  шагов. После выполнения каждого  $i$ -го шага отнесения объекта  $A_i$ ,  $i = 1, \dots, N$  к какому-либо кластеру  $O_k$ ,  $k = 1, \dots, K$ , рассчитываем значение координаты абсцисс центроидов вновь преобразованного кластера по формулам (5)–(10) или (12). Координаты оси абсцисс центроидов всех вновь преобразованных кластеров  $O_k(A_n)$  обозначим как  $G_k^j(A_n^j)$ ,  $j = 1, \dots, m$ ,  $n = 1, \dots, N$ ,  $k = 1, \dots, K$ . Переходим к п. 3.

3. Рассчитываем значения

$$w_{kn}^j = |G_k^j(A_n^j) - g_n^j|, \quad j = 1, \dots, m, \quad n = 1, \dots, N, \quad k = 1, \dots, K.$$

$$\bar{W}_{kn} = \sum_{j=1}^m \rho^j w_{kn}^j, \quad k = 1, \dots, K.$$

Объект  $A_n$  относится к кластеру  $O_{\bar{k}}$ , для которого значение  $\bar{W}_{kn}$  минимально, т.е.  $\bar{k}(A_n) = \arg \max_{1 \leq k \leq K} \bar{W}_{kn}$ . Переходим к шагу 4.

4. Если количество распределенных по кластерам объектов меньше  $N$ , то переходим к выполнению пункта 1. В противном случае алгоритм А.2 завершает свою работу.

#### Алгоритм В.

Для решения задачи 2 может быть использован любой из приведенных выше критериев оптимальности и предусматривается выполнение  $(N + K - 1)$  шагов, что заключается в следующем.

1. Выполняем вычисления пункта 1 алгоритмов А.1 и А.2. Переходим к п. 2.

2. Вычисляем значения

$$v_i = \sum_{j=1}^m g_i^j, \quad i = 1, \dots, N.$$

Упорядочим все объекты в последовательность в порядке возрастания значений  $v_i$ :

$$\tilde{U} = \{A(v_1), A(v_2), \dots, A(v_n), A(v_{n+1}), \dots, A(v_N) \mid v_n \leq v_{n+1}\}.$$

Здесь  $A(v_n)$  — объект, для которого координата оси абсцисс циклоида равна  $v_n$ . Переходим к п. 3.

3. Вычисления п. 3 предусматривают выполнение  $N$  шагов, связанных с построением кратчайшей связывающей сети системы объектов.

На каждом из этих шагов в цепь включается объект  $A_n$ , стоящий на последнем  $n$ -м месте в последовательности  $\tilde{U}$ . Обозначим кратчайшую связывающую сеть на этом этапе решения  $C(A_n)$ . Рассматривая преобразованную связывающую сеть как некоторый кластер, рассчитываем значение оси абсцисс центроида этого кластера по формулам (5)–(10) или (12). Если включены все  $N$  объектов в строящуюся цепь, то переходим к выполнению пункта 4. В противном случае продолжаем включение следующих, стоящих в последовательности, объектов.

4. На данном этапе вычислений выполняется  $(K - 1)$  шагов, связанных с разрывом участков связывающей сети в некотором выбранном  $(K - 1)$ -м месте. При этом может быть использован любой из приведенных в разделе 3 критериев оптимальности.

Пусть на некотором  $p$ -м малом шаге итерации, где  $p = 1, 2, \dots, (N - 1)$ , образовано  $(p + 1)$  кластеров, в которые включены все  $N$  рассматриваемых объектов. Отметим, что на первом шаге цепь разбивается на два кластера. Выполняем следующие вычисления:

1. На 1-м шаге итерации выбираем объект  $A_n$ , где  $n = \left\lfloor \frac{N}{2} \right\rfloor$  — целая часть от деления этих величин. Образует второй кластер. В первый кластер  $O_1$  включаем объекты  $A_i$ ,  $i = 1, 2, \dots, n$ , а во второй кластер  $O_2$  — объекты  $A_{n+1}, A_{n+2}, \dots, A_N$ . Рассчитываем координаты оси абсцисс центроидов обоих кластеров  $Z_x(O_1)$  и  $Z_x(O_2)$  по выше приведенным формулам.

2. Пусть на некотором  $p$ -м шаге, где  $p = 2, \dots, (N - 2)$  второй итерации построено  $p$  кластеров  $O_1, O_2, \dots, O_p$ . В каждом из этих кластеров содержатся подмножества объектов  $A_{k1}, A_{k2}, \dots, A_{kh}$ , которые упорядочены в последовательности убывания значений координаты оси абсцисс центроидов обоих кластеров:

$$\tilde{U}_k = \{A_{k1}, A_{k2}, \dots, A_{kh} \mid Z_x(O_k) \leq Z_x(O_{k+1}), k = 1, 2, \dots, (h - 1)\}.$$

3. В каждом из кластеров  $O_k$  выбираем соответственно два объекта  $A_{kq}$  и  $A_{k,q+1}$ , где  $q = \lfloor \frac{h}{2} \rfloor$ . Кластер  $O_k$  разбиваем на четыре кластера  $O_k^1, O_k^2, O_k^3$  и  $O_k^4$ . Каждый из этих кластеров включает следующие объекты:

$$O_k^1 = (A_{k1}, A_{k2}, \dots, A_{kq}), \quad O_k^2 = (A_{k,q+1}, A_{k,q+2}, \dots, A_{kh});$$

$$O_k^3 = (A_{k1}, A_{k2}, \dots, A_{k,q}, A_{k,q+1}), \quad O_k^4 = (A_{kq}, A_{k,q+1}, A_{k,q+2}, \dots, A_{kh}).$$

Рассчитываем значения координаты оси абсцисс центроидов обоих кластеров  $Z_x(O_k^1), Z_x(O_k^2)$  и  $Z_x(O_k^3), Z_x(O_k^4)$ . Координаты оси абсцисс центроидов всех остальных кластеров  $Z_x(O_l), l = 1, \dots, (k-1), (k+1), \dots, p$  (кроме выбранного  $k$ -го кластера) остаются без изменения.

4. Вычисляем по приведенным формулам значения одного из выбранных критериев оптимальности для каждого из вновь образованных 4-х кластеров —  $\Phi(O_k^1), \Phi(O_k^2)$  и  $\Phi(O_k^3), \Phi(O_k^4)$ . Рассматриваем также и остальные оставшиеся без изменения кластеры  $O_1, O_2, \dots, O_{k-1}, O_{k+1}, \dots, O_p$ . Для каждого разделяемого кластера  $O_k$  находим минимальное среди четырех возможных значений, которое обозначим как  $\bar{\Phi}(O_k^v)$ . Здесь  $v$  — номер выбранного кластера  $t$ , для которого достигается это минимальное значение.

Определяем индекс кластера, разбиение которого на два кластера обеспечивает минимальное значение одного из критериев оптимальности (23).

$$\bar{\Phi}_q^v = \min_{1 \leq k \leq p} \bar{\Phi}(O_k^v).$$

В зависимости от полученного решения количество кластеров увеличиваем на 1 следующим образом:

$$O_1, O_2, \dots, O_{q-1}, O_q^1, O_q^2, O_{q+1}, \dots, O_{p+1}$$

или 
$$O_1, O_2, \dots, O_{q-1}, O_q^3, O_q^4, O_{q+1}, \dots, O_{p+1}.$$

Если  $(p + 1) = K$ , то алгоритм завершает свою работу. В противном случае производим переиндексацию всех преобразованных кластеров и переходим к выполнению п. 4.

## ИЛЛЮСТРАТИВНЫЙ ПРИМЕР

Параметры треугольных функций принадлежности 2 признаков 12 объектов и 2 признаков 2 кластеров приведены в табл. 1. Значения признаков всех объектов и кластеров приведены к одним и тем же единицам измерения.

**Таблица 1.** Параметры функций принадлежности 12 объектов и 2-х кластеров

Номер объекта	Параметры функций принадлежности объектов и кластеров															
	Параметры объектов								Параметры кластеров							
	1-й признак				2-й признак				1-й признак				2-й признак			
	$a_i^1$	$c_i^1$	$b_i^1$	$x(A_i^1)$	$a_i^2$	$c_i^2$	$b_i^2$	$x(A_i^2)$	$a_k^1$	$c_k^1$	$b_k^1$	$x(O_k^1)$	$a_k^2$	$c_k^2$	$b_k^2$	$x(O_k^2)$
1	1	4	6	3,67	5	8	10	9,67	4	5	6	5,0	8	10	11	10,0
2	5	7	9	7,0	6	7	8	7,0	8	10	12	9,67	3	4	6	4,33
3	10	11	12	11,0	2	4	5	3,67								
4	3	8	9	6,67	9	10	12	10,33								
5	10	12	15	12,33	4	6	9	6,33								
6	7	8	10	11,67	2	3	4	3,0								
7	8	10	11	9,67	4	5	7	5,33								
8	3	6	9	6,0	5	8	10	7,67								
9	10	13	16	13,0	7	8	10	8,33								
10	7	10	11	9,33	1	5	6	4,0								
11	12	13	15	13,33	4	7	8	6,33								
12	6	8	10	8,0	6	7	11	7,33								

В табл. 2 приведены результаты решения иллюстративного примера алгоритмом 1 на основе применения критерия оптимальности (13).

**Таблица 2.** Результаты решения задачи алгоритмом 1

Номер объекта	Расчетные данные отнесения рассматриваемых объектов к одному из 2-х кластеров							Кластер
	Обобщенные расстояния к кластеру 1			Обобщенные расстояния к кластеру 2				
	$H_{1i}^1$	$H_{1i}^2$	$H_{1i}^1 + H_{1i}^2$	$H_{2i}^1$	$H_{2i}^2$	$H_{2i}^1 + H_{2i}^2$		
1	1,33	0,33	1,66	4,33	5,34	9,67	$O_1$	
2	2,0	3,0	5,0	2,67	2,67	5,34	$O_1$	
3	6,0	6,33	12,33	1,33	0,66	1,99	$O_2$	
4	1,67	0,33	2,0	3,0	6,0	9,0	$O_1$	
5	7,33	3,67	11,0	2,66	2,0	4,66	$O_2$	
6	6,67	7,0	13,67	2,0	1,33	3,33	$O_2$	
7	4,67	4,67	9,34	0	1,0	1,0	$O_2$	
8	1,0	2,33	3,33	3,67	3,34	7,01	$O_1$	
9	8,0	1,67	9,67	3,33	4,0	7,33	$O_2$	
10	4,33	6,0	10,33	0,34	0,33	0,67	$O_2$	
11	8,33	3,67	12,0	3,66	2,0	5,66	$O_2$	
12	3,0	2,67	5,67	1,67	3,0	4,67	$O_2$	

Следовательно,  $O_1 = \{A_1, A_2, A_4, A_8\}$ ,  $O_2 = \{A_3, A_5, A_6, A_7, A_9, A_{10}, A_{11}, A_{12}\}$ . Трапецидальные параметры функций принадлежности 2-х объектов нечетких множеств вновь образованных кластеров имеют вид:

$$\mu_{O_1^1}(O_1^1) = (1, 4, 8, 9), \mu_{O_1^2}(O_1^2) = (5, 7, 10, 12);$$

$$\mu_{O_2^1}(O_2^1) = (6, 8, 13, 16), \mu_{O_2^2}(O_2^2) = (1, 3, 8, 11),$$

а координаты абсцисс центров тяжести функций принадлежности 2-х объектов этих нечетких множеств соответственно равны:

$$Z_x[\mu_{O_1^1}(O_1^1)] = 5,068, Z_x[\mu_{O_1^2}(O_1^2)] = 8,5;$$

$$Z_x[\mu_{O_2^1}(O_2^1)] = 10,788, Z_x[\mu_{O_2^2}(O_2^2)] = 7,268.$$

Следует отметить, что при выборе второго критерия оптимальности результаты решения этой задачи могут оказаться другими.

## ЗАКЛЮЧЕНИЕ

Кластерный анализ в условиях представления многомерных объектов и кластеров нечеткими множествами имеет много практических приложений: в биологии, медицинской и технической диагностике; в проблемах охраны окружающей среды; в экономике и маркетинге, а также в системе государственного управления для классификации производителей, потребителей и ситуаций; в информатике для быстрой навигации и поиска информации в больших массивах данных и многих других областях.

Предложены алгоритмы и формульные выражения (для функций принадлежности треугольного и трапецидального вида) объединения различных многомерных объектов, параметры которых заданы fuzzy-множествами, в кластеры, вычисления координат центроидов функций принадлежности, а также расстояний между центроидами таких объектов и кластеров.

Сформулированы различные критерии оптимальности в виде минимизации средневзвешенной суммы представленной в различной метрике расстояний между центроидами объектов и кластеров, а также максимизации расстояний между центроидами различных кластеров. Рассматриваются многокритериальные подходы, основанные на аддитивной свертке различных критериев.

Рассмотрены постановки и математические формулировки, каждая из которых может использовать любой из приведенных выше критериев оптимальности. Сформулированные задачи 2 и 3 являются задачами экспоненциальной сложности в условиях большой размерности и могут быть решены только приближенными методами. Предложены полиномиальные алгоритмы получения приближенного решения двух из трех сформулированных задач. Один из алгоритмов решения задачи 1 проиллюстрирован на числовом примере. Алгоритмы решения других задач будут рассмотрены в других работах автора.

Полученные в работе результаты могут быть развиты в работах других авторов и найти применение в целом ряде практических приложений.

## ЛИТЕРАТУРА

1. J. Bacher , A. Pöge , und K. Wenzig, Clusteranalyse – *Anwendungsorientierte Einführung in Klassifikationsverfahren. 3. Auflage.* Oldenbourg, München, 2010, 538 s.
2. С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, и Л.Д.Мешалкин, *Прикладная статистика: Классификация и снижение размерности.* Москва: Финансы и статистика, 1989, 607 с.
3. И.Д. Мандель, *Кластерный анализ.* Москва: Финансы и статистика, 1988, 176 с.
4. М. Жамбю, *Иерархический кластер-анализ и соответствия.* Москва: Финансы и статистика, 1988, 345 с.
5. Ю.А. Зак, *Принятие решений в условиях размытых и нечетких данных.* Москва: URSS, 2013, 352 с.
6. Ю.А. Зак, *Принятие эффективных решений в экономике и менеджменте в условиях наличия нечисловой информации и размытых данных.* Москва: Экономика, 2018, 245 с.
7. Д.А. Вятчинин, *Нечёткие методы автоматической классификации.* Минск: Технопринт, 2004, 219 с.
8. Д.А. Вятчинин, “Применение нечетких чисел для обоснования кластеров в методах нечеткой кластеризации”, *Искусственный интеллект*, № 3, с. 523–533, 2008.
9. Ю.П. Зайченко и М.А. Гончар, “Нечеткие методы кластерного анализа в задачах автоматической классификации в экономике”, *Вісник Національного технічного університету України “КПІ”. Інформатика, управління та обчислювальна техніка*, вип. 47, с. 198–206, 2007.
10. Н.А. Гудилина, Э.С. Иванова, А.В. Сибиряков, Н.В. Мехоношина, Н.Б. Ростова, и Р.А. Файзрахманов, “Использование кластерного анализа при разработке подходов по выбору и назначению схем лечения ВИЧ-инфицированным пациентам”, *Бюллетень сибирской медицины*, № 16 (3), с. 52–60, 2017.
11. F. Höppner, F. Klawonn, R. Kruse, and T. Runkler, *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition.* Chichester: Wiley Intersciences, 1999, 289 p.
12. J.C. Bezdek, J.M. Keller, R. Krishnapuram, and N.R. Pal, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing.* New York: Springer Science, 2005, 776 p.
13. В.С. Заболотникова и О.Н. Ромашкова, “Анализ методов кластеризации для эффективного управления процессами налоговой службы”, *Фундаментальные исследования*, № 9, с. 303–330, 2017.
14. E. El-Khamy, R.A. Sadek, and M.A. El-Khoreby, “An efficient brain mass detection with adaptive clustered based fuzzy C-mean and thresholding”, *IEEE International Conference on Signal and Image Processing Applications*, pp. 429–433, 2015.
15. T. Banerjee, “Day or Night Activity Recognition From Video Using Fuzzy Clustering Techniques”, *IEEE Transactions on Fuzzy Systems*, 22 (3), pp. 483–493, 2014.
16. S. Sridevi and S. Nirmala, “ANFIS based decision support system for prenatal detection of Truncus Arteriosus congenital heart defect”, *Applied Soft Computing*, no. 46, pp. 577–587, 2016.
17. M.A.M. Abushariah, A.A.M. Alqudah, O.Y. Adwan, and R.M.M. Yousef, “Automatic Heart Disease Diagnosis System Based on Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) Approaches”, *Journal of Software Engineering and Applications*, no. 07(12), pp. 1055–1064, 2014.
18. S. Sridevi and S. Nirmala, “ANFIS based decision support system for prenatal detection of Truncus Arteriosus congenital heart defect”, *Applied Soft Computing*, 46, pp. 577–587, 2016.
19. Н.А. Кореновский, “Использование нечеткой логики принятия решений для медицинских экспертных систем”, *Медицинская техника*, № 1, с. 33–35, 2015.

Поступила 01.05.2021

## INFORMATION ON THE ARTICLE

**Yuriy A. Zack**, Dr. Tech. Sc., scientific expert, Deutschland, e-mail: yuriy\_zack@hotmail.com

### КЛАСТЕРНИЙ АНАЛІЗ ДЛЯ БАГАТОВИМІРНИХ ОБ'ЄКТІВ В УМОВАХ НЕЧІТКИХ ДАНИХ / Ю.О. Зак

**Анотація.** Наведено описані в літературі безлічі різних галузей практичних застосувань багатомірних кластерного аналізу в умовах нечітких вихідних даних. Запропоновано нові алгоритми і формульні вирази об'єднання різних багатомірних об'єктів, параметри яких задано fuzzy-множинами, у кластери і обчислення координат центрів їх функцій належності. Сформульовано різні види критеріїв кластеризації у вигляді мінімізації середньозваженої і поданої в різній метриці суми відстаней між центрами об'єктів і кластерів, а також максимізації відстаней між центрами різних кластерів. Запропоновано постановки і математичні моделі трьох різних NP-складних задач багатомірної кластеризації в умовах fuzzy-даних, для розв'язання яких може бути використаний будь-який з розглянутих критеріїв оптимальності. Розроблено евристичні алгоритми наближеного розв'язку двох сформульованих задач. Алгоритм розв'язання першої задачі проілюстровано на числовому прикладі. Отримані результати можуть стати напрямом подальших досліджень і знайти широке практичне застосування.

**Ключові слова:** кластерний аналіз, багатомірні функції належності, центри fuzzy-множин об'єктів і кластерів, центри ваги і середини перетинів нечітких множин, критерії оптимальності та алгоритми кластеризації.

### CLUSTER ANALYSIS FOR MULTIDIMENSIONAL OBJECTS IN FUZZY DATA CONDITIONS / Yu.A. Zack

**Abstract.** This article presents many different areas of practical applications of multivariate cluster analysis under conditions of fuzzy initial data that are described in the literature. New algorithms and formula expressions are proposed for combining various multi-dimensional objects, the parameters of which are given by fuzzy-sets, into clusters along with calculating the coordinates of the centroids of their membership functions. Various types of clustering criteria are formulated in the form of minimizing the weighted average and the sum of distances between the centroids of objects and clusters presented in different metrics, as well as maximizing the distances between the centroids of different clusters. The formulations and mathematical models of three different NP-hard problems of multidimensional clustering in fuzzy-data conditions are proposed; while solving them any of the considered optimality criteria can be used. Heuristic algorithms for the approximate solution of two formulated problems have been developed. The algorithm for solving the 1st problem is illustrated with a numerical example. The obtained results can serve as a direction for further research and have wide practical applications.

**Keywords:** cluster analysis, multidimensional membership functions, centroids of fuzzy-sets of objects and clusters, centers of gravity and mid-sections of fuzzy sets, optimality criteria and clustering algorithms.

### REFERENCES

1. J. Bacher, A. Pöge, K. Wenzig, *Clusteranalyse – Anwendungsorientierte Einführung in Klassifikationsverfahren*. 3. Auflage. Oldenbourg, München, 2010, 538 s.
2. S.A. Ajvazjan, V.M. Buchstaber, I.S. Enyukov, and L.D. Meshalkin, *Applied Statistics: Classification and Dimension Reduction*. Moscow: Finansi & Statistika, 1989, 607 p.
3. I.D. Mandel, *Cluster Analysis*. Moscow: Finansi & Statistika, 1988, 176 p.
4. M.Zhambyu, *Hierarchical Cluster Analysis and Compliance*. Moscow: Finansi & Statistika, 1988, 345 p.



5. Yu.A. Zack, *Making decisions in the face of blurry and fuzzy data*. Moscow: URSS, 2013, 352 p.
6. Yu.A. Zack, *Making effective decisions in economics and management in the presence of non-numerical information and blurry data*. Moscow: Ekonomika, 2018, 245 p.
7. D.A. Vyatchenin, *Fuzzy automatic classification methods*. Minsk: Technoprint, 2004, 219 p.
8. D.A. Vyatchenin, "The use of fuzzy numbers to substantiate clusters in fuzzy clustering methods", *Iskusstvenij intellect*, no. 3, pp. 523–533, 2008.
9. Yu.P. Zaychenko and M.A. Gonchar, "Fuzzy methods of cluster analysis in problems of automatic classification in economics", *Bulletin of the National Technical University of Ukraine "KPI". Informatics, control and computer engineering*, vol. 47, pp. 198–206, 2007.
10. N.A. Grudinina, E.S. Ivanova, A.V. Sibiryakov, N.V. Maüjnshina, N.B. Rostova, R.A. Fajsrachmanov, "The use of cluster analysis in the development of approaches to the selection and prescription of treatment regimens for HIV-infected patients", *Bulletin of Siberian Medicine*, no. 16 (3), pp. 52–60, 2017.
11. F. Höppner, F. Klawonn, R. Kruse, and T. Runkler, *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition*. Chichester: Wiley Intersciences, 1999, 289 p.
12. J.C. Bezdek, J.M. Keller, R. Krishnapuram, and N.R. Pal, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*. New York: Springer Science, 2005, 776 p.
13. V.S. Zabolotnikova and O.N. Romashkova, "Analysis of clustering methods for effective management of tax service processes", *Fundamental Research*, no. 9, pp. 303–330, 2017.
14. E. El-Khamy, R.A. Sadek, M.A. El-Khoreby, "An efficient brain mass detection with adaptive clustered based fuzzy C-mean and thresholding", *IEEE International Conference on Signal and Image Processing Applications*, pp. 429–433, 2015.
15. T. Banerjee, "Day or Night Activity Recognition From Video Using Fuzzy Clustering Techniques", *IEEE Transactions on Fuzzy Systems*, 22 (3), pp. 483–493, 2014.
16. S. Sridevi and S. Nirmala, "ANFIS based decision support system for prenatal detection of Truncus Arteriosus congenital heart defect", *Applied Soft Computing*, no. 46, pp. 577–587, 2016.
17. M.A.M. Abushariah, A.A.M. Alqudah, O.Y. Adwan, and R.M.M. Yousef, "Automatic Heart Disease Diagnosis System Based on Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) Approaches", *Journal of Software Engineering and Applications*, no. 07(12), pp. 1055–1064, 2014.
18. S. Sridevi and S. Nirmala, "ANFIS based decision support system for prenatal detection of Truncus Arteriosus congenital heart defect", *Applied Soft Computing*, 46, pp. 577–587, 2016.
19. N.A. Korenovsky, "The use of fuzzy decision-making logic for medical expert systems", *Medical equipment*, no. 1, pp. 33–35, 2015.