

УДК 519.677

## Алгоритмы кластерного анализа для решения задач с асимметричной мерой близости\*

А.Р. Айдинян, О.Л. Цветкова

Донской государственный технический университет, пл. Гагарина, 1, Ростов-на-Дону, 344000

E-mails: aydinian.andrey@yandex.ru (Айдинян А.Р.), olga\_cvetskova@mail.ru (Цветкова О.Л.)

**Айдинян А.Р., Цветкова О.Л.** Алгоритмы кластерного анализа для решения задач с асимметричной мерой близости // Сиб. журн. вычисл. математики / РАН. Сиб. отд-ние. — Новосибирск, 2018. — Т. 21, № 2. — С. 127–138.

Кластерный анализ используется в различных научных и прикладных областях и является актуальной темой исследований. В отличие от существующих методов предложенные в работе алгоритмы предназначены для кластеризации объектов, описываемых векторами признаков в пространстве с несоблюдением аксиомы симметрии. В этом случае особенностью решения задачи кластеризации является использование асимметричной меры близости объектов.

Суть первого из предложенных алгоритмов кластеризации заключается в последовательном формировании кластеров с одновременным перенесением кластеризованных объектов из ранее созданных кластеров в текущий кластер в случае, если это уменьшит критерий качества. По сравнению с существующими алгоритмами неиерархической кластеризации такой подход к формированию кластеров позволяет уменьшить вычислительные затраты. Второй алгоритм является модифицированной версией первого и дополнительно позволяет выполнять переназначения главных объектов кластера с целью дальнейшего уменьшения величины предложенного критерия качества.

**DOI:** 10.15372/SJNM20180201

**Ключевые слова:** кластеризация, кластерный анализ, алгоритмы кластеризации, асимметричная мера близости, аксиома симметрии.

**Aydinyan A.R., Tsvetkova O.L.** The cluster algorithms for solving problems with asymmetric proximity measures // Siberian J. Num. Math. / Sib. Branch of Russ. Acad. of Sci. — Novosibirsk, 2018. — Vol. 21, № 2. — P. 127–138.

The cluster analysis is used in various fundamental and applied fields and is a current topic of research. Unlike conventional methods, the proposed algorithms are used for clustering objects represented by vectors in space with the non-observance of the axiom of symmetry. In this case, the feature of solving the clustering problem is the use of an asymmetric proximity measures.

The first one among the proposed clustering algorithms sequentially forms clusters with a simultaneous generalization to clustered objects from previously created clusters to a current cluster if this reduces the quality criterion. This approach to the formation of clusters allows reducing the computational costs as compared with existing non-hierarchical cluster algorithms. The second algorithm is a modified version of the first algorithm. The second algorithm allows reassigning the main objects of clusters to further reduce the proposed quality criterion.

**Keywords:** clustering, cluster analysis, cluster algorithms, asymmetric proximity measure, the axiom of symmetry.

---

## 1. Введение

Кластерный анализ представляет собой задачу разбиения (кластеризации) заданной выборки объектов на некоторое количество подмножеств — кластеров, таким образом,

---

\*Работа выполнена при финансовой поддержке Российского научного фонда (проект № 14-11-0083).

чтобы каждый кластер содержал в себе объекты, которые более схожи между собой, чем с объектами из других кластеров [1]. Схожесть объектов может выражаться различными способами, и для ее определения предлагаются различные меры близости [2].

Обзор опубликованных исследований [3, 4] позволяет сделать заключение о том, что в настоящее время кластерный анализ нашел широкое применение в таких областях как медицина, археология, социология, маркетинг, статистика, распознавание образов, оптимизация, финансовая математика, выработка стратегий управления, исследование свойств ДНК и других [5–7]. В [8–10] решается важная для многих естественно-научных и технических приложений задача классификации упорядоченных по времени результатов численных экспериментов. К задаче кластеризации может быть отнесена задача распределения работ между несколькими исполнителями [11].

Особенностью решения задач кластерного анализа является то, что в большинстве случаев приходится использовать несколько алгоритмов или перебирать параметры прежде, чем будет получен приемлемый результат [12].

Общепринятой классификации методов кластеризации не существует, например, интересные подходы к классификации представлены в работах [2, 13, 14]. Рисунок 1 иллюстрирует результаты анализа различных подходов к классификации методов кластеризации. Все методы кластеризации авторы предлагают разделить по способу обработки данных на иерархические и неиерархические и по способу анализа данных на четкие и нечеткие.

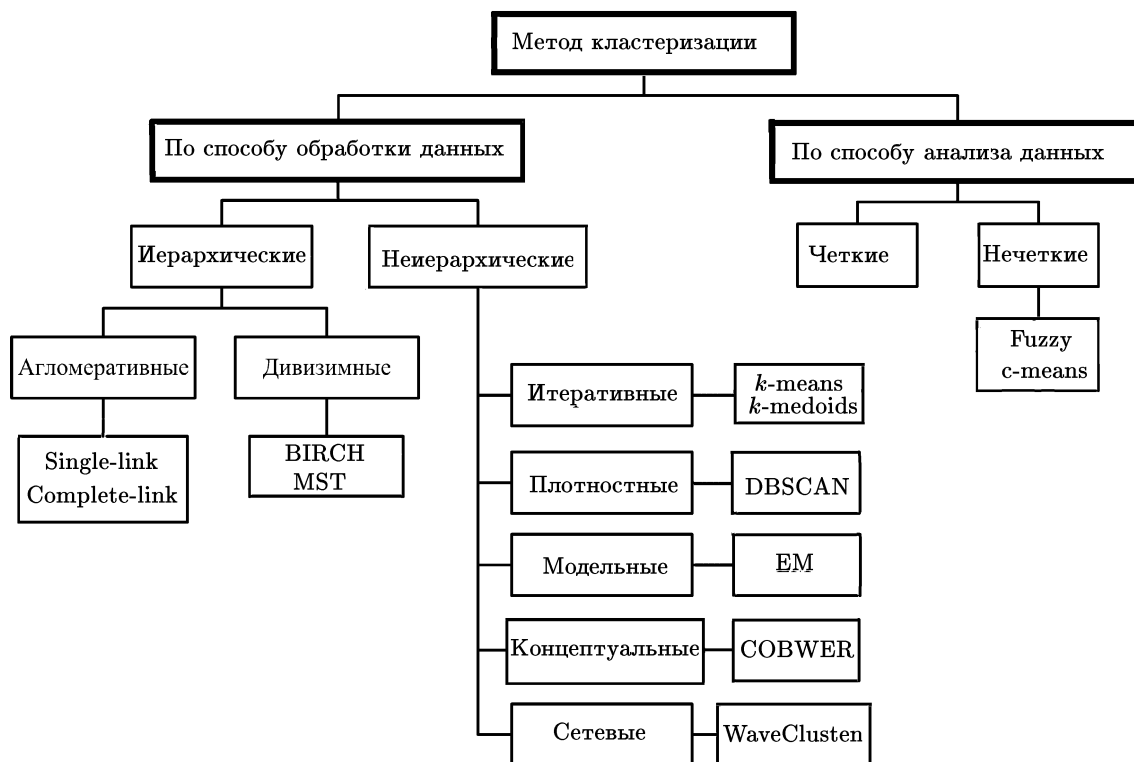


Рис. 1. Классификация методов кластеризации

Рассмотренные методы кластеризации обладают определенными достоинствами и недостатками, которые обеспечивают их эффективное применение при решении конкретных задач. Общим для этих методов является то, что они используют меру близости объектов в метрическом пространстве, для которого расстояние между объектами  $A$ ,  $B$

и  $C$  должно удовлетворять следующим аксиомам:

$d(A, B) > d(A, A) = 0$  — минимальность;

$d(A, B) = d(B, A)$  — симметрия;

$d(A, B) + d(B, C) \geq d(A, C)$  — неравенство треугольника.

Однако имеется ряд задач, в которых необходимо разделить некоторое количество объектов на кластеры с выделением одного из них в качестве главного с учетом того, что расстояние от объекта  $A$  к объекту  $B$  не равно расстоянию от объекта  $B$  к объекту  $A$ .

В качестве прикладных задач, в которых возникает необходимость кластеризации объектов при условии несимметричности функции расстояния, можно предложить следующие:

- 1) разбиение территории, обслуживаемой некоторой торговой компанией, на участки, в каждом из которых находится главный склад и торговые точки. В случае если компания имеет участки земли с известными координатами и расстоянием между ними, то задача заключается в разбиении всех этих участков на кластеры с выбором местоположения одного главного склада в каждом из кластеров. При этом расстояние между участками в одном и другом направлениях может быть разным. Под расстоянием может пониматься функция затрат, зависящая от расходов топлива, времени движения, платы за движение по платным дорогам и др. Асимметричность в данном случае может возникать в связи с наличием пробок или запретом движения в одном из направлений в определенное время суток, наличием платной дороги в одном направлении, наличием дорог с односторонним движением, запретом въезда некоторых видов транспорта в определенном направлении;
- 2) выбор страны, в которой производится продукция для экспорта. В этом случае асимметричность возникает в связи с разными таможенными пошлинами в разные страны, налоговыми сборами, стоимостью рабочей силы, стоимостью перевозки готовой продукции;
- 3) выбор центральных (центральной) точек при протекании физических процессов. Например, выбор расположения точек, откуда будет распространяться жидкость в другие точки с целью минимизации затрат энергии на ее перемещение. Асимметричность в этом случае возникает в связи с разной потенциальной энергией жидкости в разных точках.

В статье предлагаются четкие алгоритмы кластеризации при условии асимметричной меры близости между кластеризуемыми объектами. Эти алгоритмы относятся к группе неиерархических методов, но не требуют априорного задания количества кластеров.

## 2. Постановка задачи

При решении задачи кластеризации исходными данными является множество объектов  $W$ , каждый из которых описывается набором характеристик, называемых признаками  $W = \{X_1, \dots, X_n\}$ . Каждому  $i$ -му объекту ставится в соответствие вектор признаков  $X_i = (x_{i,1}, \dots, x_{i,m})$ ,  $i = 1, \dots, n$ . Элемент  $x_{i,j}$  ( $j = 1, \dots, m$ ) содержит  $j$ -й признак  $i$ -го объекта. Количество признаков  $m$  определяет размерность пространства признаков объекта.

Формальная постановка задачи четкой кластеризации заключается в том, что на основании данных, содержащихся в векторах  $X_i$ ,  $i = 1, \dots, n$ , необходимо разбить множество объектов  $W$  на  $K$  кластеров так, чтобы каждый объект принадлежал только одному подмножеству разбиения, и объекты, принадлежащие одному и тому же кластеру, были “близкими”, а объекты, принадлежащие разным кластерам, существенно отличались.

Мера близости объектов задается с помощью функции  $d(X_i, X_j)$ ,  $i, j = 1, \dots, n$ , определяющей расстояние между объектами  $X_i$  и  $X_j$  [2, 15–17]. При этом в существующих работах предполагается, что  $d(X_i, X_j) = d(X_j, X_i)$ .

В данной статье предлагаются методы кластеризации объектов с признаками, представленными в пространстве с несоблюдающейся аксиомой симметрии, т. е. возможно  $d(X_i, X_j) \neq d(X_j, X_i)$  при  $i \neq j$ . Подобные задачи кластеризации возникают, когда в каждом  $k$ -м кластере необходимо выделить главный объект  $\bar{X}_k$ , по сути являющийся медоидом [18], таким образом, чтобы обеспечивался минимум суммы

$$\sum_{X_i \in V_k} d(\bar{X}_k, X_i),$$

где  $k$  — рассматриваемый кластер;  $V_k$  — множество объектов  $k$ -го кластера. При этом нет необходимости минимизировать  $\sum_{X_i \in V_k} d(X_i, \bar{X}_k)$ .

### 3. Описание предлагаемых алгоритмов кластеризации

Предлагаемые алгоритмы кластеризации предназначены для эффективного решения задач, когда признаки объектов расположены в пространстве с асимметричной мерой близости. Алгоритмы различаются по точности кластеризации, временным затратам и затратам ресурсов памяти.

Предлагаемые алгоритмы стремятся разбить множество объектов  $W$  на  $K$  кластеров, причем значение  $K$  перед выполнением алгоритмов не задается и может косвенно регулироваться выбором максимально допустимого радиуса кластера  $R_{\max}$  таким образом, чтобы достичь минимизации критерия

$$F = \sum_{k=1}^K \sum_{i=1}^{r_k} d(\bar{X}_k, X_i^k) \rightarrow \min,$$

где  $r_k$  — количество объектов, отнесенных к  $k$ -му кластеру;  $\bar{X}_k$  — главный объект  $k$ -го кластера;  $X_i^k$  —  $i$ -й объект  $k$ -го кластера.

Ясно, что в общем случае значение рассмотренного критерия будет равно нулю при равенстве количества кластеров количеству объектов. Однако такое решение является неприемлемым в практических приложениях, поскольку имеются расходы на создание кластера или главного объекта кластера. Поэтому предлагается ввести в критерий слабое, зависящее от количества кластеров

$$F = \sum_{k=1}^K \sum_{i=1}^{r_k} d(\bar{X}_k, X_i^k) + K S \rightarrow \min,$$

или

$$F = \sum_{k=1}^K \left( \sum_{i=1}^{r_k} d(\bar{X}_k, X_i^k) \right) + \sum_{k=1}^K S_k \rightarrow \min,$$

где  $S$  — средние расходы на организацию одного кластера;  $S_k$  — расходы на организацию  $k$ -го кластера.

Величина  $S$  позволяет учесть в критерии качества количество кластеров, что является существенным при решении задач, в которых затраты прямо пропорциональны количеству созданных кластеров. Например, в задаче разбиения территории, обслуживаемой

торговой компаний, и выбора расположения складов имеются расходы на организацию складов. Величина  $S$  в предложенных алгоритмах учитывается косвенно только на этапе выбора величины  $R_{\max}$ . В случае отсутствия  $S$  оптимальным решением будет размещение каждого объекта в отдельный кластер (для минимизации необходимо будет выбрать число кластеров, равным числу объектов, поместив в каждый кластер по одному объекту).

Исходными данными для предлагаемых алгоритмов являются значения функции  $d(X_i, X_j)$ ,  $i, j = 1, \dots, n$ , и максимально допустимый радиус кластеров  $R_{\max}$ . Первоначально все объекты  $X_i$  принадлежат множеству некластеризованных объектов  $W = \{X_i\}_{i=1}^n$ . Множества  $V_k$ ,  $k = 1, \dots, K$ , для хранения объектов каждого  $k$ -го кластера создаются по мере необходимости.

Новизна предложенных алгоритмов заключается в возможности решения задач кластеризации с асимметричной мерой близости. Алгоритмы основываются на неиерархической кластеризации и позволяют уменьшить вычислительные затраты за счет последовательного формирования кластеров с сопутствующим перенесением в текущий кластер ранее кластеризованных объектов в случае, если это приводит к уменьшению критерия качества  $F$ , а точнее, его первого слагаемого (величина  $S$  учитывается косвенно на этапе выбора величины  $R_{\max}$ ). Итоговая величина критерия качества  $F$  зависит также от количества созданных кластеров и влияет на принимаемое решение о приемлемости полученного результата.

**Алгоритм 1.** Алгоритм с возможностью переноса кластеризованных объектов в текущий кластер.

**Шаг 1.** Инициализация алгоритма. Присвоить номеру текущего кластера 1 ( $k = 1$ ).

**Шаг 2.** Выбор главного объекта текущего кластера. За главный объект  $k$ -го кластера принять объект  $\bar{X}_k$ :

$$\bar{X}_k = \arg \min_x \{d(x, y) \mid x, y \in W, x \neq y\},$$

исключить его из множества  $W$ :

$$W = W \setminus \{\bar{X}_k\}$$

и поместить в  $k$ -й кластер (добавить в множество  $V_k$ , предназначенное для хранения объектов  $k$ -го кластера):

$$V_k = \{\bar{X}_k\}.$$

**Шаг 3.** Добавление объектов в текущий кластер. Добавить в  $k$ -й кластер некластеризованные объекты, расстояние до которых от главного объекта кластера не больше  $R_{\max}$ :

$$V_k = V_k \cup \{w \in W \mid d(\bar{X}_k, w) \leq R_{\max}\}$$

и исключить эти объекты из множества  $W$ :

$$W = W \setminus \{V_k\}.$$

**Шаг 4.** Перенесение объектов в текущий кластер. В созданных ранее кластерах найти множество объектов  $M$ , состоящее из объектов  $m$ , для которых расстояние до них от главного объекта  $k$ -го кластера  $\bar{X}_k$  меньше, чем расстояние до них от главных объектов кластеров, в котором эти объекты находятся:

$$M = \{m \in V_p \mid d(\bar{X}_k, m) < d(\bar{X}_p, m), p = 1, \dots, k-1\}.$$

Множество найденных объектов  $M$  исключить из соответствующих множеств  $V_p$  и перенести в  $k$ -й кластер:

$$V_p = V_p \setminus M, \quad p = 1, \dots, k-1,$$

$$V_k = V_k \cup M.$$

**Шаг 5.** Расформирование кластеров. Для всех  $j$ -х кластеров ( $j = 1, \dots, k-1$ ) с одним объектом  $\bar{X}_j$  перенести объекты  $\bar{X}_j$  в существующие  $r$ -е кластеры, для которых расстояние от главного объекта  $\bar{X}_r$  до объекта  $\bar{X}_j$  является минимальным:

$$r = \arg \min_p d(\bar{X}_p, \bar{X}_j), \quad \bar{X}_p \neq \bar{X}_j, \quad p = 1, \dots, k,$$

$$V_r = V_r \cup \{\bar{X}_j\},$$

$$V_j = \{\}.$$

**Шаг 6.** Переход к следующему кластеру. Если имеются некластеризованные объекты, то увеличить номер текущего кластера на 1 ( $k = k+1$ ) и перейти к шагу 2. Иначе завершить алгоритм.

Второй разработанный алгоритм является модифицированной версией алгоритма 1 и дополнительно позволяет выполнять переназначения главных объектов кластеров в случае, если это может привести к уменьшению критерия качества.

**Алгоритм 2.** Алгоритм с возможностью переназначения главных объектов кластеров и переноса кластеризованных объектов в текущий кластер.

**Шаг 1.** Инициализация алгоритма (см. шаг 1 алгоритма 1).

**Шаг 2.** Выбор главного объекта текущего кластера (см. шаг 2 алгоритма 1).

**Шаг 3.** Добавление объектов в текущий кластер (см. шаг 3 алгоритма 1).

**Шаг 4.** Выбор нового главного объекта текущего кластера. Последовательным перебором найти объект  $x \in V_k$ , который обеспечит минимальное значение суммы расстояний от него до всех объектов текущего  $k$ -го кластера,

$$z = \arg \min_x \left\{ \sum_{y \in V_k} d(x, y) \mid \forall x \in V_k \right\}.$$

Если  $z \neq \bar{X}_k$ , то принять его в качестве главного:

$$\bar{X}_k = z.$$

**Шаг 5.** Если на шаге 4 был переназначен главный объект  $k$ -го кластера, то перейти к шагу 3.

**Шаг 6.** Перенесение кластеризованных объектов в текущий кластер (см. шаг 4 алгоритма 1).

**Шаг 7.** Расформирование кластеров (см. шаг 5 алгоритма 1).

**Шаг 8.** Переход к следующему кластеру (см. шаг 6 алгоритма 1).

Вычислительная сложность предложенных алгоритмов равна  $O(n^2)$ , где  $n$  — число кластеризуемых объектов.

Предложенные алгоритмы обладают следующими особенностями:

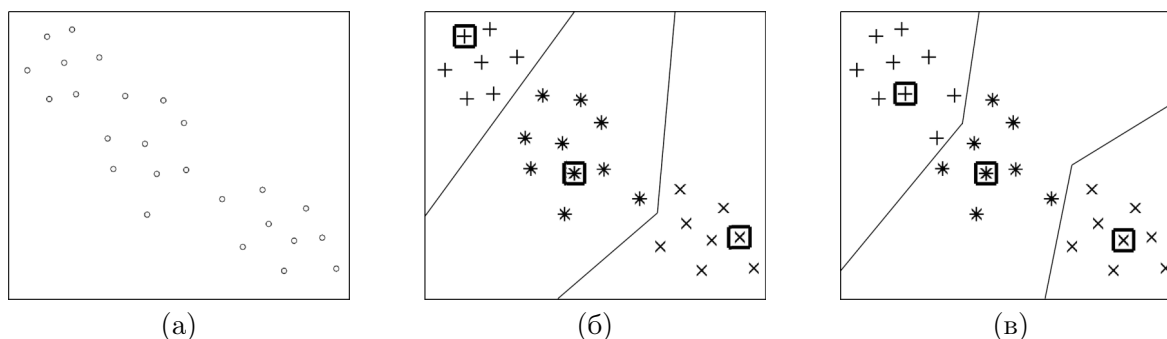
- 1) при создании кластера выбираются два объекта:  $x, y \in W$  такие, что расстояние  $d(x, y) \rightarrow \min$ , при этом объект  $x$  принимается за главный объект кластера;
- 2) выполняется последовательное создание кластеров, что позволяет уменьшить вычислительные затраты;
- 3) кластер в алгоритме 2 может пополняться объектами как из множества некластеризованных, так и из множеств ранее кластеризованных объектов с целью минимизации критерия;
- 4) переназначение главного объекта кластера осуществляется с целью дальнейшего уменьшения критерия качества без переформирования кластера. Это приводит к тому, что расстояние от главного объекта до остальных объектов кластера может превышать заданную величину  $R_{\max}$ .

#### 4. Программная реализация алгоритмов и анализ результатов

Реализация алгоритмов кластеризации была выполнена с помощью пакета MatLab 2014b, который содержит средства для определения времени вычислений.

Экспериментальные исследования предложенных алгоритмов кластеризации проводились для нескольких наборов исходных данных, отличающихся количеством и параметрами объектов, подлежащих кластеризации, и значением максимального радиуса кластеров  $R_{\max}$ .

**Пример 1.** Кластеризация 25 объектов при  $R_{\max} = 0.8$  и  $R_{\max} = 1.2$ . Выборка, используемая в примере 1, состоит из координат земельных участков, которыми владеет организация. Расстояния между объектами являются асимметричными и представляют собой затраты на перемещения между участками. Результаты кластеризации приведены на рис. 2 и в таблице 1.



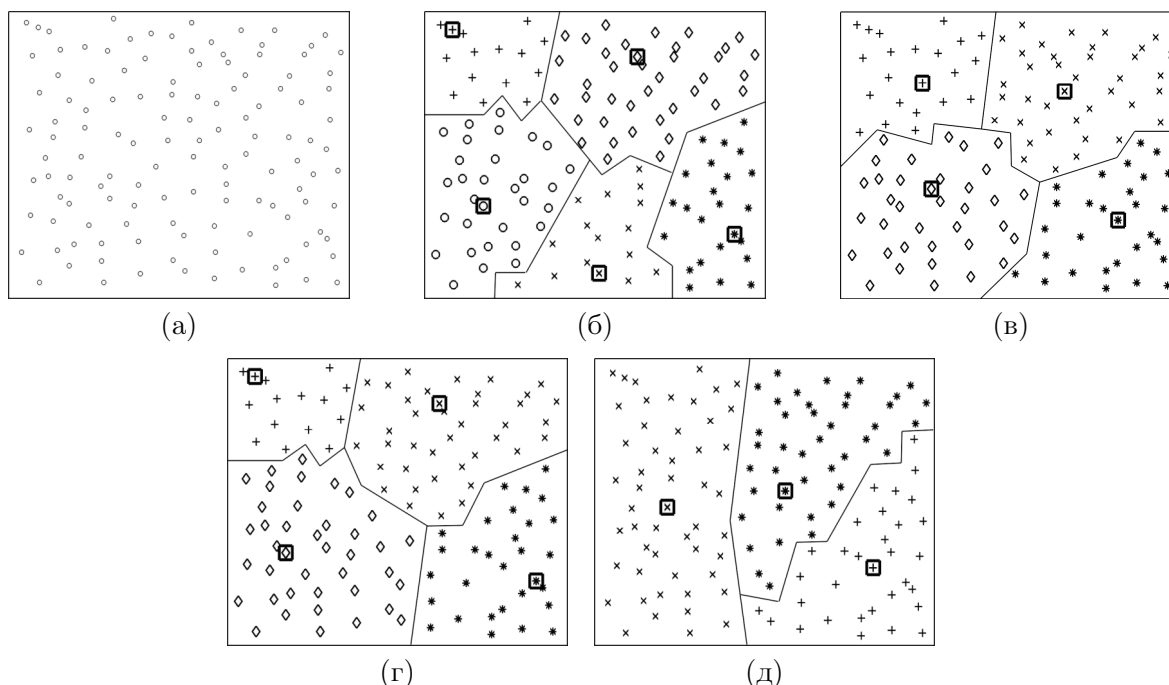
**Рис. 2.** Графическое отображение кластеров, сформированных алгоритмами в результате кластеризации 25 объектов при  $R_{\max} = 0.8$ : (а) — расположение объектов на плоскости; (б), (в) — кластеризация алгоритмами 1, 2

**Таблица 1.** Результаты кластеризации 25 объектов

| Параметр                                                   | Алгоритм 1,<br>$R_{\max} = 0.8$ | Алгоритм 2,<br>$R_{\max} = 0.8$ | Алгоритм 1,<br>$R_{\max} = 1.2$ | Алгоритм 2,<br>$R_{\max} = 1.2$ |
|------------------------------------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| Количество объектов в кластерах 1/2/3                      | 7/10/8                          | 9/8/8                           | 7/10/8                          | 16/9/—                          |
| Оценка кластера —<br>сумма расстояний<br>в кластерах 1/2/3 | 2.637/<br>4.663/<br>2.692       | 3.691/<br>3.295/<br>2.692       | 2.434/<br>4.255/<br>3.100       | 9.595/<br>3.197/<br>—           |
| Оценка решения —<br>сумма расстояний для всех кластеров    | 9.992                           | 9.678                           | 9.789                           | 12.792                          |
| Время работы алгоритма, с                                  | 0.017166                        | 0.022282                        | 0.01074                         | 0.01865                         |

**Пример 2.** Кластеризация 125 объектов с программно сгенерированными асимметричными расстояниями между ними с равномерным распределением.

На рис. 3 и в табл. 2 приведены результаты кластеризации при  $R_{\max}=1.2$  и  $R_{\max}=1.7$ .



**Рис. 3.** Графическое отображение кластеров, сформированных предложенными алгоритмами в результате кластеризации 125 объектов: (а) — расположение объектов на плоскости; (б), (в) — кластеризация алгоритмами 1, 2 при  $R_{\max} = 1.2$ ; (г), (д) — кластеризация алгоритмами 1, 2 при  $R_{\max} = 1.7$

**Таблица 2.** Результаты кластеризации 125 объектов

| Параметр                                                       | Алгоритм 1,<br>$R_{\max} = 1.2$          | Алгоритм 2,<br>$R_{\max} = 1.2$    | Алгоритм 1,<br>$R_{\max} = 1.7$   | Алгоритм 2,<br>$R_{\max} = 1.7$ |
|----------------------------------------------------------------|------------------------------------------|------------------------------------|-----------------------------------|---------------------------------|
| Количество объектов<br>в кластерах 1/2/3/4/5                   | 15/24/16/39/31                           | 21/29/38/36/—                      | 15/30/40/40/—                     | 32/42/51/—/—                    |
| Оценка кластера —<br>сумма расстояний<br>в кластерах 1/2/3/4/5 | 8.626/13.603/<br>9.226/26.163/<br>17.637 | 11.257/16.625/<br>23.383/22.900/ — | 8.626/19.881/<br>27.313/27.399/ — | 20.453/34.213/<br>42.234/ — / — |
| Оценка решения —<br>сумма расстояний<br>для всех кластеров     | 75.255                                   | 74.165                             | 83.219                            | 96.900                          |
| Время работы, с                                                | 0.22414                                  | 0.22314                            | 0.21107                           | 0.20988                         |

**Пример 3.** Кластеризация выборки Iris Data Set из репозитория UCI. В репозитории UCI нет выборок с асимметричными мерами близости, для работы с которыми предназначены разработанные алгоритмы. Тем не менее предложенные алгоритмы можно использовать также для кластеризации объектов, имеющих признаки, заданные в пространстве с соблюдением аксиомы симметрии. Разработанные алгоритмы имеют большее сходство с алгоритмом  $k$ -medoids. Поэтому сравним результаты, полученные алгоритмом 2 и алгоритмом  $k$ -medoids, по критерию минимизации суммы расстояний для всех кластеров.

Выборка Iris Data Set состоит из 150 элементов трех классов. Каждый класс представляет различный сорт цветов ириса. В каждом классе представлено по 50 элементов. Поскольку сорт ириса зависит от ширины и длины лепестка, из выборки были отброшены параметры, определяющие размер чашелистика, и затем отброшены повторяющиеся значения параметров лепестка. В результате в выборке осталось 102 элемента, которые подвергались кластеризации. Расстояние между элементами выборки вычислялось как квадрат евклидова расстояния.

В табл. 3 приведены результаты сравнительного анализа разработанного алгоритма 2 и алгоритма  $k$ -medoids по сумме квадратов евклидовых расстояний между медоидами и остальными объектами кластеров при различном количестве сгенерированных кластеров.

**Таблица 3.** Результаты сравнительного анализа алгоритма 2 с алгоритмом  $k$ -medoids на выборке Iris Data Set из репозитория UCI

| Алгоритм     | Оценка решения — сумма расстояний для всех кластеров |          |          |          |          |           |
|--------------|------------------------------------------------------|----------|----------|----------|----------|-----------|
|              | Количество кластеров / $R_{\max}$                    |          |          |          |          |           |
|              | 2 / 6.00                                             | 3 / 2.10 | 4 / 0.90 | 6 / 0.58 | 8 / 0.38 | 10 / 0.20 |
| Алгоритм 2   | 72.3325                                              | 26.264   | 17.22    | 9.4646   | 6.9513   | 5.04      |
| $k$ -medoids | 72.36                                                | 26.20    | 16.63    | 9.33     | 6.73     | 4.91      |

Результаты кластеризации, полученные в примерах, показывают:

- алгоритм 1 позволяет получить решение при значительной экономии временных затрат по сравнению с алгоритмом 2 в случае, если количество сформированных кластеров одинаковое;
- алгоритм 2 при одинаковой величине  $R_{\max}$  формирует меньшее количество кластеров по сравнению с алгоритмом 1;
- алгоритм 2 по сравнению с алгоритмом 1 позволяет получить меньшую величину суммы расстояний внутри всех кластеров при больших вычислительных затратах в случае формирования одинакового количества кластеров;
- алгоритм 2 при одинаковой величине  $R_{\max}$  в большинстве случаев генерирует меньшее количество кластеров при большей величине критерия качества практически без изменения временных затрат;
- при использовании алгоритма 2 для получения такого же количества кластеров, как и алгоритм 1, обычно необходимо уменьшить величину  $R_{\max}$ . При этом величина критерия качества уменьшится, а временные затраты увеличатся;
- алгоритм 2 при использовании симметричных расстояний между объектами позволяет получить результаты близкие к результатам, полученным алгоритмом  $k$ -medoids, в пространстве с соблюдением аксиомы симметрии и сравнении их по

параметру суммарного расстояния от главных объектов кластеров (медоид) до остальных объектов кластеров при одинаковом количестве кластеров.

Авторы могут дать следующие рекомендации по использованию разработанных алгоритмов кластеризации. Алгоритм 1 является наиболее эффективным в случаях, когда требуется получить достаточно точное решение с экономией временных ресурсов. Если требуется получить меньшую величину критерия качества без учета временных затрат при том же значении  $R_{\max}$ , то следует применять алгоритм 2.

## Заключение

Разработаны алгоритмы кластерного анализа, обеспечивающие выполнение кластеризации объектов, векторы признаков которых расположены в пространстве с асимметричной мерой близости. Предложенные алгоритмы отличаются степенью достижения заданного критерия качества и вычислительными затратами при одинаковой величине  $R_{\max}$ . Актуальность предложенных алгоритмов связана с необходимостью решения подобного рода задач в различных областях научной и практической деятельности.

## Литература

1. **Tryon R.C.** Cluster analysis. — London: Ann Arbor Edwards Bros, 1939.
2. **Madhulatha T.S.** An overview on clustering methods // IOSR J. of Engineering. — 2012. — Vol. 2, № 4. — P. 719–725.
3. **Hartigan J.A.** Clustering Algorithms (Probability & Mathematical Statistics). — New York – London – Sydney – Toronto: John Wiley & Sons Inc, 1975.
4. **Sörensen T.** A method of establishing groups of equal amplitude in plant sociology based on similarity of species content // Kongelige Danske Videnskabernes Selskab. Biol. Skrifter. — 1948. — Bd. V, № 4. — P. 1–34.
5. **Safari K.M., Tive M., Babania A., and Hesani M.** Analyzing the applications of customer lifetime value (CLV) based on benefit segmentation for the banking sector // Procedia — Social and Behavioral Sciences. 2nd World Conference on Business, Economics and Management. — 2014. — Vol. 109. — P. 590–594.
6. **Srihadi D.T.F., Hartoyo, Sukandar D., and Soehadi A.W.** Segmentation of the tourism market for Jakarta: Classification of foreign visitors' lifestyle typologies // Tourism Management Perspectives. — 2016. — Vol. 19, № 1. — P. 32–39.
7. **Beh A., Bruyere B.L.** Segmentation by visitor motivation in three Kenyan national reserves // Tourism Management. — 2007. — Vol. 28. — P. 1464–1471.
8. **Кельманов А.В., Хамидуллин С.А.** Приближенный полиномиальный алгоритм для одной задачи бикластеризации последовательности // Журн. вычисл. матем. и мат. физики. — 2015. — Т. 55, № 6. — С. 1076–1085. — Перевод: Kel'manov A.V., Khamidullin S.A. An approximation polynomial-time algorithm for a sequence BI-clustering problem // Computational Mathematics and Mathematical Physics. — 2015. — Vol. 55, № 6. — P. 1068–1076.
9. **Кельманов А.В., Романченко С.М., Хамидуллин С.А.** Точные псевдополиномиальные алгоритмы для некоторых труднорешаемых задач поиска подпоследовательности векторов // Журн. вычисл. матем. и мат. физики. — 2013. — Т. 53, № 1. — С. 143–153.
10. **Kel'manov A.V., Jeon B.A.** Posteriori joint detection and discrimination of pulses in a quasiperiodic pulse train // IEEE Trans. on Signal Proc. — 2004. — Vol. 52, № 3. — P. 645–656.

11. **Айдинян А.Р., Цветкова О.Л.** Генетические алгоритмы распределения работ // Вестник Донского государственного технического университета. — 2011. — Т. 11, № 5 (56). — С. 723–729.
12. **Шалымов Д.С.** Алгоритмы устойчивой кластеризации на основе индексных функций и функций устойчивости // В сб. Стохастическая оптимизация в информатике / О.Н. Гранича. — СПб.: Изд-во С.-Петербургского университета. — 2008. — Вып. 4. — С. 236–248.
13. **Jain A.K., Murty M.N., and Flynn P.J.** Data clustering: a review // ACM Computing Surveys. — 1999. — Vol. 31, № 3. — P. 264–323.
14. **Soni N., Ganatra A.** Categorization of several clustering algorithms from different perspective: a review // Int. J. of Advanced Research in Computer Science and Software Engineering. — 2012. — Vol. 2, № 8. — P. 63–68.
15. **Mao J., Jain A.K.** A self-organizing network for hyperellipsoidal clustering (HEC) // IEEE Trans. Neural Netw. — 1996. — Vol. 7. — P. 16–29.
16. **Huttenlocher D.P., Klanderman G.A., and Rucklidge W.J.** Comparing images using the Hausdorff distance // IEEE Trans. Pattern Anal. Mach. Intell. — 1993. — Vol. 15, № 9. — P. 850–863.
17. **Dubuisson M.P., Jain A.K.** A modified Hausdorff distance for object matching // Proc. of the International Conference on Pattern Recognition (ICPR '94). — 1994. — P. 566–568.
18. **Van der Lann M.J., Pollard K.S., and Bryan J.** A new partitioning around medoids algorithm // J. of Statistical Computation and Simulation. — 2003. — Vol. 5, № 8. — P. 575–584.

*Поступила в редакцию 22 ноября 2016 г.,  
в окончательном варианте 9 сентября 2017 г.*

## Литература в транслитерации

1. **Tryon R.C.** Cluster analysis. — London: Ann Arbor Edwards Bros, 1939.
2. **Madhulatha T.S.** An overview on clustering methods // IOSR J. of Engineering. — 2012. — Vol. 2, № 4. — P. 719–725.
3. **Hartigan J.A.** Clustering Algorithms (Probability & Mathematical Statistics). — New York – London – Sydney – Toronto: John Wiley & Sons Inc, 1975.
4. **Sörensen T.** A method of establishing groups of equal amplitude in plant sociology based on similarity of species content // Kongelige Danske Videnskabernes Selskab. Biol. Skrifter. — 1948. — Bd. V, № 4. — P. 1–34.
5. **Safari K.M., Tive M., Babania A., and Hesani M.** Analyzing the applications of customer lifetime value (CLV) based on benefit segmentation for the banking sector // Procedia — Social and Behavioral Sciences. 2nd World Conference on Business, Economics and Management. — 2014. — Vol. 109. — P. 590–594.
6. **Srihadi D.T.F., Hartoyo, Sukandar D., and Soehadi A.W.** Segmentation of the tourism market for Jakarta: Classification of foreign visitors' lifestyle typologies // Tourism Management Perspectives. — 2016. — Vol. 19, № 1. — P. 32–39.
7. **Beh A., Bruyere B.L.** Segmentation by visitor motivation in three Kenyan national reserves // Tourism Management. — 2007. — Vol. 28. — P. 1464–1471.
8. **Kel'manov A.V., Hamidullin S.A.** Priblizhennyj polinomial'nyj algoritm dlya odnoj zadachi biklasterizacii posledovatel'nosti // Zhurn. vychisl. matem. i mat. fiziki. — 2015. — Т. 55, № 6. — С. 1076–1085. — Перевод: Kel'manov A.V., Khamidullin S.A. An approximation polynomial-time algorithm for a sequence BI-clustering problem // Computational Mathematics and Mathematical Physics. — 2015. — Vol. 55, № 6. — P. 1068–1076.

9. **Kel'manov A.V., Romanchenko S.M., Hamidullin S.A.** Tochnye psevdopolynomial'nye algoritmy dlya nekotoryh trudnoreshaemyh zadach poiska podposledovatel'nosti vektorov // Zhurn. vychisl. matem. i mat. fiziki. — 2013. — Т. 53, № 1. — С. 143–153.
10. **Kel'manov A.V., Jeon B.A.** Posteriori joint detection and discrimination of pulses in a quasiperiodic pulse train // IEEE Trans. on Signal Proc. — 2004. — Vol. 52, № 3. — P. 645–656.
11. **Ajdinyan A.R., Cvetkova O.L.** Geneticheskie algoritmy raspredeleniya rabot // Vestnik Donskogo gosudarstvennogo tekhnicheskogo universiteta. — 2011. — Т. 11, № 5 (56). — С. 723–729.
12. **Shalymov D.S.** Algoritmy ustojchivoj klasterizacii na osnove indeksnyh funkcij i funkcij ustojchivosti // V sb. Stohasticheskaya optimizaciya v informatike / O.N. Granichipa. — SPb.: Izd-vo S.-Peterburgskogo universiteta. — 2008. — Vyp. 4. — С. 236–248.
13. **Jain A.K., Murty M.N., and Flynn P.J.** Data clustering: a review // ACM Computing Surveys. — 1999. — Vol. 31, № 3. — P. 264–323.
14. **Soni N., Ganatra A.** Categorization of several clustering algorithms from different perspective: a review // Int. J. of Advanced Research in Computer Science and Software Engineering. — 2012. — Vol. 2, № 8. — P. 63–68.
15. **Mao J., Jain A.K.** A self-organizing network for hyperellipsoidal clustering (HEC) // IEEE Trans. Neural Netw. — 1996. — Vol. 7. — P. 16–29.
16. **Huttenlocher D.P., Klanderman G.A., and Rucklidge W.J.** Comparing images using the Hausdorff distance // IEEE Trans. Pattern Anal. Mach. Intell. — 1993. — Vol. 15, № 9. — P. 850–863.
17. **Dubuisson M.P., Jain A.K.** A modified Hausdorff distance for object matching // Proc. of the International Conference on Pattern Recognition (ICPR '94). — 1994. — P. 566–568.
18. **Van der Lann M.J., Pollard K.S., and Bryan J.** A new partitioning around medoids algorithm // J. of Statistical Computation and Simulation. — 2003. — Vol. 5, № 8. — P. 575–584.