КЛАСТЕРНЫЙ АНАЛИЗ И ЕГО СВЯЗЬ С БИЛИНЕЙНЫМ ПРОГРАММИРОВАНИЕМ*

А. Р. Баранцев

alexei.barantsev@gmail.com

16 марта 2023 г.

Кластерный анализ является одним из важнейших методов первичной обработки данных [3]. Он состоит в нахождении естественного в некотором смысле разделения множества объектов на определенное число подмножеств (кластеров). Известно, что эта задача многоэкстремальна. Существует большое число методов нахождения локально экстремальных разбиений [2]. Глобально оптимальная кластеризация считается значительно более сложной задачей. Более того, насколько известно автору, попыток детерминистского определения глобально оптимального разбиения пока не предпринималось.

В данной статье рассматривается в качестве иллюстративного примера задача разбиения на кластеры фиксированных масс. Эту задачу удается поставить как задачу билинейного программирования, тесно связанную с задачей выпуклой максимизации. Для поиска глобального экстремума задачи билинейного программирования разработан ряд алгоритмов [12]. В свою очередь, эти методы пока не применялись в кластерном анализе [14]. Цель данной статьи — установить связь между этими двумя независимо развивающимися направлениями.

Рассмотренная задача представляет также самостоятельный интерес (для случая равных масс кластеров — см. [3, гл. 3].

В статье использована индексная техника, предложенная И. В. Романовским [1].

1°. Постановка задачи. Пусть N, P — конечные множества индексов, $|P| < |N|$. Разбиение множества N на $|P|$ непересекающихся подмножеств будем рассматривать как однозначное отображение $\varphi : N \to P$. Такое отображение удобно задавать матрицей $\sigma[N, P]$, состоящей из нулей и единиц:

$$\sigma[n,p] = \begin{cases} 1, & \text{если } \varphi(n) = p, \\ 0, & \text{если } \varphi(n) \neq p. \end{cases}$$

*Семинар по оптимизации, машинному обучению и искусственному интеллекту «O&ML» `http://oml.cmlaboratory.com/`

Таким образом, в каждой строке $\sigma[n, P]$, $n \in N$, есть ровно один ненулевой элемент; он равен единице.

Матрицу $\varepsilon[N, P]$, удовлетворяющую условиям

$$\varepsilon[N, P] \geqslant \mathbf{0}[N, P], \tag{1}$$

$$\varepsilon[N, P] \times \mathbf{1}[P] = \mathbf{1}[N], \tag{2}$$

будем называть *матрицей вероятностного распределения.*

Очевидно, что любая целочисленная матрица вероятностного распределения ε задает некоторое разбиение. Любое разбиение $\varphi : N \to P$ можно представить матрицей $\varepsilon[N, P]$ вида (1)–(2) с целыми элементами.

2°. Пусть K — конечное множество индексов; $g[K]$ — набор положительных чисел. Рассмотрим пространство $\mathcal{X}$ строк вида $\xi[K]$ со скалярным произведением, определяемым метрическими коэффициентами g,

$$\langle \xi, \eta \rangle = \sum_{k \in K} g[k]\xi[k]\eta[k],$$

и нормой

$$\|\xi\| = \sqrt{\langle \xi, \xi \rangle}.$$

Пусть $x[N, K]$ — некоторая матрица. Будем ее рассматривать как семейство $|N|$ элементов пространства $\mathcal{X}$:

$$x[n, \cdot] \in \mathcal{X}, \quad n \in N.$$

Строки матрицы $x[N, K]$ будут называться также *точками*, *векторами.*

Каждой точке $x[n, \cdot]$, $n \in N$, сопоставим положительное число $\mu[n]$, которое назовем его *массой.* Пары $\{x[n, \cdot], \mu[n]\}$, $n \in N$, будем называть *объектами.*

Совокупность объектов с индексами из множества $\varphi^{-1}(p)$, $p \in P$, называется *кластером.* Другими словами, кластер с индексом $p \in P$ есть множество объектов с индексами $n \in N$, такими, что $\sigma[n, p] = 1$. Масса этого кластера равна, очевидно, $\mu[N] \times \sigma[N, p]$. Назовем массой кластерного индекса p выражение $\mu[N] \times \varepsilon[N, p]$, где ε — некоторая матрица вероятностного распределения.

Будем рассматривать задачу поиска матрицы вероятностного распределения объектов $\{x[N, \cdot], \mu[N]\}$ при заданных массах кластерных индексов. Именно, наложим на матрицу ε ограничение

$$\mu[N] \times \varepsilon[N, P] = m[P], \tag{3}$$

где $m[P]$ — некоторый набор положительных чисел. Подразумевается, что наборы m и μ согласованы,

$$m[P] \times \mathbf{1}[P] = \mu[N] \times \mathbf{1}[N]. \tag{4}$$

3°. Задача кластерного анализа состоит в нахождении наиболее естественного в некотором смысле разбиения семейства объектов $\{x, \mu\}$. Часто для этого используют минимизацию дисперсионного критерия.

Пусть $y[p, \cdot]$, $p \in P$, — некоторый набор точек пространства $\mathcal{X}$. Назовем *дисперсионным критерием* функцию f матриц $y[P, K]$, $\varepsilon[N, P]$:

$$f(y, \varepsilon) := \sum_{p \in P} \sum_{n \in N} \mu[n] \varepsilon[n, p] \|x[n, \cdot] - y[p, \cdot]\|^2. \tag{5}$$

Точки $y[p, \cdot]$ называются *представителями кластерных индексов* $p \in P$.

Если ε — целочисленная матрица, $\varepsilon = \sigma$, то дисперсионный критерий f представляет собой сумму *моментов инерции* объектов $\{x, \mu\}$ относительно представителей кластеров, к которым они отнесены.

Естественно искать распределение и набор представителей, минимизирующие дисперсионный критерий f. Будем называть *f-задачей* задачу минимизации $f(y, \varepsilon)$ (5) при ограничениях (1)–(3).

Матрицы y, ε есть неизвестные f-задачи; наборы μ, m, g, x, удовлетворяющие условиям:

$$\begin{aligned} &\mu[N] > \mathbf{0}[N], && m[P] > \mathbf{0}[P], \\ &g[K] > \mathbf{0}[K], && x[n, K] \in \mathcal{X}, \quad \forall n \in N, \end{aligned}$$

а также *условию согласования* (4), есть параметры f-задачи.

Сформулированная задача есть *задача кластеризации при фиксированных массах кластеров*, расширенная в том смысле, что искомая матрица вероятностного распределения ε может, вообще говоря, оказаться нецелочисленной, $\varepsilon \neq \sigma$.

4°. Выделим следующий набор представителей:

$$c[p, K] := \frac{1}{m[p]} \sum_{n \in N} \varepsilon[n, p] \mu[n] x[n, K]; \quad p \in P.$$

Обозначим точкой, «$\cdot$» , операцию покомпонентного умножения. С ее помощью последнюю формулу можно переписать в эквивалентном виде

$$m[P] \cdot c[P, K] = \big(\varepsilon^\top[P, N] \cdot \mu[N]\big) \times x[N, K]. \tag{6}$$

Будем называть точки $c[p, \cdot]$ *центрами тяжести кластерных индексов* $p \in P$ для вероятностного распределения ε. Если матрица ε целочисленная, то центры тяжести кластерных индексов есть центры тяжести кластеров.

Рассмотрим матрицу

$$\delta[P, N] := \frac{\varepsilon^\top[P, N] \cdot \mu[N]}{\varepsilon^\top[P, N] \times \mu[N]}. \tag{7}$$

Как следует из формул (6), (3), матрица δ представляет собой набор коэффициентов разложения координат центров тяжести $c[P,K]$ по строкам матрицы $x[N,K]$:

$$c[P,K] = \delta[P,N] \times x[N,K]. \tag{8}$$

Нетрудно убедиться в выполнении соотношений

$$\begin{aligned} \delta[P,N] &\geqslant \mathbf{0}[P,N], && (9)\\ \delta[P,N] \times \mathbf{1}[N] &= \mathbf{1}[P], && (10)\\ m[P] \cdot \delta[P,N] &= \varepsilon^{\top}[P,N] \cdot \mu[N], && (11)\\ m[P] \times \delta[P,N] &= \mu[N]. && (12) \end{aligned}$$

Чтобы получить последнюю формулу, надо просуммировать (11) по $p \in P$ и воспользоваться соотношением (2).

Отметим связь между формулами (11) и (6). Если (11) покомпонентно умножить на $x[N,K]$ и просуммировать по $n \in N$, то, в силу (8), получится соотношение (6).

5°. Следующее утверждение устанавливает разрешимость f-задачи.

УТВЕРЖДЕНИЕ 1. *Оптимум f-задачи существует. Среди оптимальных решений f-задачи найдется пара $\{y, \varepsilon\}$ такая, что*

$$m[P] \cdot y[P,K] = \left(\varepsilon^{\top}[P,N] \cdot \mu[N]\right) \times x[N,K], \tag{13}$$

то есть, $y[P,K] = c[P,K]$ есть набор центров тяжести для распределения ε.

Условие (13) будем называть *условием центровки* для пары $\{y, \varepsilon\}$.

Доказательство. Пусть пара $\{y, \varepsilon\}$ — некоторая допустимая пара f-задачи. Зафиксируем матрицу вероятностного распределения $\varepsilon[N,P]$. Тогда функция f выпукла в области изменения переменных y, как следует непосредственно из ее определения (5). На переменные $y[P,K]$ не накладывается ограничений. Стало быть, необходимое и достаточное условие оптимальности матрицы y при фиксированных ε имеет вид

$$\frac{\partial f(y,\varepsilon)}{\partial y[P,K]} = \mathbf{0}[P,K].$$

Эта система линейных уравнений эквивалентна условию центровки (13). Таким образом, $f(c,\varepsilon) \leqslant f(y,\varepsilon)$ для любой матрицы y.

Далее, в силу соотношений (8)–(10), каждый из векторов $c[p,K]$, $p \in P$, принадлежит выпуклой оболочке векторов $x[n,K]$, $n \in N$. Следовательно, с точки зрения оптимума f-задачи, можно считать, что векторы y находятся в выпуклой оболочке векторов x.

Поскольку переменные ε также принадлежат ограниченному замкнутому множеству, а целевая функция f непрерывна, то минимум f-задачи существует, что и требовалось доказать. □

Отметим, что установленное экстремальное свойство центров тяжести справедливо только в пространстве с диагональной метрикой.

6°. Дополнительная функция Отметим сходство между формулами (1)–(3) и (9), (10), (12). Это сходство будет более полным, если рассмотреть матрицы

$$\begin{aligned} e[N,P] &:= \mu[N]\cdot\varepsilon[N,P], && (14)\\ d[P,N] &:= m[P]\cdot\delta[P,N]. && (15)\end{aligned}$$

Тогда соотношения (1)–(3) и (9)–(12) принимают вид:

$$\begin{aligned} e[N,P] &\geqslant \mathbf{0}[N,P], && (16)\\ e[N,P]\times\mathbf{1}[P] &= \mu[N], && (17)\\ \mathbf{1}[N]\times e[N,P] &= m[P], && (18)\\ d[P,N] &\geqslant \mathbf{0}[P,N], && (19)\\ d[P,N]\times\mathbf{1}[N] &= m[P], && (20)\\ \mathbf{1}[P]\times d[P,N] &= \mu[N], && (21)\\ d^{\top}[N,P] &= e[N,P]. && (22)\end{aligned}$$

Формулы (16)–(18) превращаются в формулы (19)–(21) при транспонировании (22). Последнее соотношение, как было указано в п. 4°, гарантирует выполнение условия центровки (6).

Будем называть равенство (22) *условием центровки* для пары $\{d,e\}$.

7°. Пусть J — конечное множество индексов; $\varepsilon_j[N,P]$, $j\in J$, — набор матриц вероятностного распределения с заданными массами кластерных индексов $m[P]$. Построим соответствующие им матрицы $\delta_j[P,N]$, $c_j[P,K]$, $e_j[N,P]$, $d_j[P,N]$ по формулам (7), (8), (14), (15).

Пусть $i,j\in J$. Рассмотрим дисперсионный критерий f (5) на аргументах $c_i[P,K]$, $\varepsilon_j[N,P]$:

$$\begin{aligned} f(c_i,\varepsilon_j) &= \sum_{p\in P}\sum_{n\in N}\mu[n]\varepsilon_j[n,p]\langle x[n,\cdot]-c_i[p,\cdot],x[n,\cdot]-c_i[p,\cdot]\rangle =\\ &= \sum_{p\in P}\sum_{n\in N}\mu[n]\varepsilon_j[n,p]\langle x[n,\cdot],x[n,\cdot]\rangle -\\ &\quad -2\sum_{p\in P}\sum_{n\in N}\mu[n]\varepsilon_j[n,p]\langle x[n,\cdot],c_i[p,\cdot]\rangle +\\ &\quad +\sum_{p\in P}\sum_{n\in N}\mu[n]\varepsilon_j[n,p]\langle c_i[p,\cdot],c_i[p,\cdot]\rangle =:\\ &=: \sum\nolimits_{XX}-2\sum\nolimits_{CX}+\sum\nolimits_{CC}.\end{aligned}$$

В сумме $\sum_{XX}$ поменяем порядок суммирования по $p \in P$ и по $n \in N$. Воспользовавшись условием (2), получаем

$$\sum_{XX} = \sum_{n \in N} \mu[n] \langle x[n,\cdot], x[n,\cdot] \rangle =: M_0.$$

Величина M_0 есть полный момент инерции системы точек $x[N,\cdot]$ относительно начала координат; M_0 не зависит от переменных f-задачи.

Сумму $\sum_{CX}$ выразим через матрицы d_i, e_j, воспользовавшись формулами (14), (8), а затем (15):

$$\begin{aligned}\sum_{CX} &= \sum_{p \in P} \sum_{n \in N} e_j[n,p] \sum_{n' \in N} \langle x[n,\cdot], x[n',\cdot] \rangle \delta_i[p,n'] = \\ &= \sum_{p \in P} \sum_{n \in N} \sum_{n' \in N} e_j[n,p] \frac{\langle x[n,\cdot], x[n',\cdot] \rangle}{m[p]} d_i[p,n'] = \\ &= \sum_{p \in P} \frac{d_i[p,N] \times R[N,N] \times e_j[N,p]}{m[p]},\end{aligned}$$

где через R обозначена матрица Грама системы векторов $x[N,\cdot]$,

$$R[n,n'] := \langle x[n,\cdot], x[n',\cdot] \rangle; \quad n, n' \in N.$$

Назовем *дополнительной функцией* следующую функцию аргументов $d[P,N]$, $e[N,P]$:

$$F(d,e) := \sum_{p \in P} \frac{d[p,N] \times R[N,N] \times e[N,p]}{m[p]}. \tag{23}$$

Таким образом, $\sum_{CX} = F(d_i, e_j)$.

При вычислении суммы $\sum_{CC}$ воспользуемся выражением (6) для центров тяжести кластерных индексов:

$$\begin{aligned}\sum_{CC} &= \sum_{p \in P} m[p] \langle c_i[p,\cdot], c_i[p,\cdot] \rangle = \\ &= \sum_{p \in P} \sum_{n \in N} \mu[n] \varepsilon_i[n,p] \langle x[n,\cdot], c_i[p,\cdot] \rangle.\end{aligned}$$

Эта сумма отличается от $\sum_{CX}$ только тем, что при $\varepsilon[n,p]$ стоит индекс i вместо j. Стало быть, $\sum_{CC} = F(d_i, e_i)$.

Итак, имеет место соотношение

$$f(c_i, \varepsilon_j) = M_0 - 2F(d_i, e_j) + F(d_i, e_i),$$

где пара $\{d_i, e_j\}$ связана с парой $\{c_i, \varepsilon_j\}$ посредством формул (14), (15), (7), (8).

Отметим, что указанная связь между f и F справедлива только при фиксировании масс кластерных индексов. Условие (3) существенно использовано при вычислении $\sum_{CC}$.

8°. Введем матрицы

$$f[i,j] := f(c_i, \varepsilon_j), \qquad F[i,j] := F(d_i, e_j); \quad i,j \in J.$$

Ниже всегда будет ясно, где f или F используется для обозначения функции, а где — для обозначения матрицы. Индексы множества J иногда будет удобнее писать внизу: f_{ij}, F_{ij}.

Матрицы $f[J,J]$, $F[J,J]$ назовем *матрицами значений дисперсионного критерия и дополнительной функции* соответственно.

Как установлено в предыдущем пункте, введенные матрицы связаны соотношением

$$f_{ij} = M_0 - 2F_{ij} + F_{ii}. \tag{24}$$

В частности, при $i = j$ имеем

$$f_{ii} = M_0 - F_{ii}. \tag{25}$$

Последняя формула выражает теорему Гюйгенса из механики, согласно которой сумма внутри- и межклассовой инерции есть полная инерция [3, гл. 14].

Нетрудно получить из (24), (25) формулы

$$f_{ij} - f_{ii} = -2(F_{ij} - F_{ii}), \tag{26}$$

$$f_{ij} - f_{jj} = F_{ii} - 2F_{ij} + F_{jj}, \tag{27}$$

$$F_{ij} = M_0 - \frac{1}{2}(f_{ij} + f_{ii}). \tag{28}$$

Таким образом, если зафиксировать некоторый i-й набор центров тяжести (соответственно, матрицу d_i), то изменения f и F при изменении распределения противоположны по знаку (формула (26)).

Приращение дисперсионного критерия при изменении центров тяжести (первый индекс) определяется формулой (27).

Формула (28) выражает F через f; это обращение формулы (24).

9°. Отметим ряд свойств дополнительной функции F и ее матрицы значений $F[J,J]$.

УТВЕРЖДЕНИЕ 2. *Матрица значений $F[J,J]$ симметрична и неотрицательно определена.*

Доказательство. Матрица Грама $R[N,N]$ симметрична. Пользуясь условием центровки (22) для пары $\{d,e\}$, получаем

$$\begin{aligned} F[i,j] &= \sum_{p\in P}\sum_{n\in N}\sum_{n'\in N} \frac{d_i[p,n]R[n,n']e_j[n',p]}{m[p]} = \\ &= \sum_{p\in P}\sum_{n\in N}\sum_{n'\in N} \frac{e_i^\top[p,n]R[n',n]d_j^\top[n',p]}{m[p]} = \end{aligned}$$

$$\begin{aligned} &= \sum_{p\in P}\sum_{n\in N}\sum_{n'\in N}\frac{d_j[p,n']R[n',n]e_i[n,p]}{m[p]} = \\ &= F[j,i] = F^\top[i,j]. \end{aligned}$$

Чтобы доказать неотрицательную определенность матрицы F, рассмотрим квадратичную форму

$$\begin{aligned} \Phi(\lambda) &:= \lambda[J]\times F[J,J]\times\lambda[J] = \\ &= \sum_{p\in P}\frac{1}{m[p]}\sum_{i\in J}\sum_{j\in J}\lambda_i e_i^\top[p,N]\times R[N,N]\times e_j[N,p]\lambda_j. \end{aligned}$$

Обозначим $e_\lambda[N,P] := \sum_{j\in J}\lambda_j e_j[N,P]$. Тогда

$$\Phi(\lambda) = \sum_{p\in P}\frac{1}{m[p]}e_\lambda^\top[p,N]\times R[N,N]\times e_\lambda[N,p] = F(e_\lambda^\top, e_\lambda).$$

Поскольку матрица Грама неотрицательно определена, а числа $m[p]$, положительны, то каждое слагаемое в последней сумме неотрицательно. Следовательно,

$$\forall\lambda[J]\quad \Phi(\lambda) = F(e_\lambda^\top, e_\lambda) = \lambda[J]\times F[J,J]\times\lambda[J] \geqslant 0, \tag{29}$$

что и требовалось. □

Если в (29) положить $J := \{i,j\}$, $\lambda_i := 1$, $\lambda_j := -1$, то получим полезное неравенство

$$F[i,i] + F[j,j] \geqslant 2F[i,j]. \tag{30}$$

Отметим связь (30) с экстремальным свойством центров тяжести (утверждение 1). Из соотношений (30), (25) следует $f_{jj} \leqslant f_{ij}$ или $f(c_j,\varepsilon_j) \leqslant f(c_i,\varepsilon_j)$.

Кроме того, неравенство (30) позволяет применить теорию билинейного программирования ([4], [5], [7], [8], [9], [15]) к решению задачи кластерного анализа.

Введем квадратичную функцию

$$Q(e) := F(e^\top, e) = \sum_{p\in P}\frac{e^\top[p,N]\times R[N,N]\times e[N,p]}{m[p]}. \tag{31}$$

Очевидно, что $Q(e)\geqslant 0$ при всех e. Если в (29) положить $J := \{i,j\}$, $\lambda_i := 1$, $\lambda_j := 1$, то получим

$$Q(e_i + e_j) = Q(e_i) + 2F(e_i^\top, e_j) + Q(e_j).$$

В частности, $Q(e_i{+}e_j){-}Q(e_i) \geqslant 2F(e_i^\top, e_j)$. Последнее неравенство гарантирует выпуклость Q ([1, гл. 3, лемма 3.1, достаточность]).

10°. Максимизация дополнительной функции. Назовем *F-задачей* следующую задачу: максимизировать $F(d, e)$ (23) при ограничениях (16)–(21).

Матрицы d и e есть неизвестные F-задачи. Целевая функция F билинейна, ограничения линейны. Ограничения распадаются на две группы: (16)–(18) относятся к переменным e, а (19)–(21) — к переменным d. Задача симметрична относительно замены $d \to e^\top$, $e \to d^\top$.

Матрица R и наборы m, μ — параметры F-задачи. Матрица $R[N, N]$ симметрична и неотрицательно определена; наборы m и μ положительны и согласованы (см. (4)).

Отметим, что условие центровки (22) не входит в число ограничений F-задачи.

Наряду с F-задачей рассмотрим связанную с ней Q-задачу: максимизировать $Q(e) = F(e^\top, e)$ при ограничениях (16)–(18).

Она получается из F-задачи при наложении условия центровки (22) на пару $\{d, e\}$.

11°. Согласно (30), значение дополнительной функции на некоторой паре $\{d_i, e_j\}$ не превосходит среднего арифметического значений F на соответствующих центрованных парах. В частности, это означает, что либо $F(d_i, d_i^\top) \geqslant F(d_i, e_j)$, либо $F(e_j^\top, e_j) \geqslant F(d_i, e_j)$. Благодаря этому обстоятельству, справедливы следующие два утверждения, доказательство которых имеется в [7].

УТВЕРЖДЕНИЕ 3. *Если пара $\{d, e\}$ локально оптимальна для F-задачи, то пары $\{d, d^\top\}$, $\{e^\top, e\}$, $\{e^\top, d^\top\}$ — тоже локально оптимальны; при этом F имеет одинаковое значение на всех этих парах. Более того, тогда $d^\top$ и e — локально оптимальные решения Q-задачи.*

УТВЕРЖДЕНИЕ 4. *Если e — локальный максимум для Q-задачи, то пара $\{e^\top, e\}$ доставляет локальный максимум F-задаче.*

Следующее утверждение аналогично утверждению 1 для f-задачи:

УТВЕРЖДЕНИЕ 5. *F-задача разрешима. Среди ее оптимальных решений найдется пара $\{d, e\}$, удовлетворяющая условию центровки (22).*

Доказательство. Разрешимость F-задачи следует из непрерывности целевой функции и ограниченности и замкнутости допустимой области. Вторая часть утверждения очевидным образом вытекает из утверждения 3. $\square$

12°. Теперь можно показать, что f-задача сводится к билинейной F-задаче.

ТЕОРЕМА. *Пусть пара $\{d, e\}$ — оптимальное решение F-задачи, удовлетворяющая условию центровки (22). Тогда пара $\{c, \varepsilon\}$, полученная из $\{d, e\}$ с помощью (14), (15), (8), есть оптимальное решение f-задачи.*

Доказательство. В силу утверждения 5, пара $\{d, e\}$, удовлетворяющая условиям теоремы, существует. Согласно утверждению 3, матрица e решает Q-задачу. При выполнении условий центровки справедлива теорема Гюйгенса (25): сумма f и F постоянна. Тогда соответствующая пара $\{c, \varepsilon\}$ минимизирует f при условии центровки (13). Эта пара есть оптимальное решение f-задачи вследствие утверждения 1, что и требовалось. $\square$

Итак, задача кластерного анализа при заданных массах кластерных индексов, поставленная как задача минимизации дисперсионного критерия f в пространстве $\mathcal{X}$ с диагональной метрической матрицей, сведена к F-задаче максимизации билинейной функции F на выпуклом многограннике с ограничениями транспортного типа.

Такие задачи, как правило, обладают неединственным локально оптимальным решением. В кластерном анализе обычно ограничиваются нахождением локальных экстремумов; глобальная оптимизация считается столь сложной, что, насколько известно автору, попыток ее решить не предпринималось [2]. С другой стороны, для решения задачи билинейного программирования разработан ряд методов [12], которые можно применить для решения задачи кластеризации.

Задача кластеризации при фиксированных массах кластеров имеет ограниченную область применения (см. [3, гл. 3 при равных массах кластеров]). Однако, в отношении вида целевой функции, эта задача оказывается простейшей из задач кластерного анализа.

Для нахождения локального экстремума этой задачи применяется метод мультикритериальной классификации [3, гл. 3]. Локальные экстремумы задачи билинейного программирования находят поочередной оптимизацией по каждой переменной [9]. Такой алгоритм, однако, может привести не к локальному экстремуму, а к седловой точке, находящейся в вершине допустимого многогранника. Ниже описан аналогичный алгоритм для F-задачи.

13°. Система ограничений (16)–(18) F-задачи на матрицу e есть система ограничений транспортного типа. Матрица $e[N, P]$ может быть интерпретирована как матрица перевозок; наборы m и μ — как векторы потребления и производства. Если зафиксировать матрицу d, то матрицу

$$\frac{d[P, N] \times R[N, N]}{m[P]}$$

с индексами из P, N можно интерпретировать как матрицу расценок.

Таким образом, F-задача при фиксированных $d[N, P]$ становится транспортной задачей, в которой целевая функция максимизируется, а матрица расценок может не быть неотрицательной. Последнее обстоятельство, однако, не изменяет алгоритма ее решения.

То же самое можно сказать и о F-задаче при фиксированных $e[N, P]$.

Стационарную точку F-задачи в смысле [1, гл. 3] можно найти с помощью следующего алгоритма.

Шаг 0. Найти начальное решение e_0 для ограничений (16)–(18) любым из способов, используемых при решении транспортной задачи. Транспонировав $e_0[N,P]$, получим матрицу $d_0[P,N]$, удовлетворяющую ограничениям (19)–(21). Пара $\{d_0, e_0\}$ допустима для F-задачи и удовлетворяет условию центровки (23). Присвоить переменной цикла j нулевое значение. Инициализировать индексное множество $J := \{0\}$. Вычислить $F_{00} = F(d_0, e_0)$.

Шаг 1. Имеется допустимая пара $\{d_j, e_j\}$, удовлетворяющая условию центровки (23). Зафиксировать матрицу d_j и поставить транспортную F-задачу с переменной матрицей e. Если e_j — ее оптимальное решение, то пара $\{d_j, e_j\}$ — стационарная точка F-задачи [1, гл. 3]. Алгоритм окончен. В противном случае перейти к шагу 2.

Шаг 2. Решить транспортную задачу при фиксированных d_j. Пусть $e_{j+1}[N,P]$ — ее оптимальное базисное решение. Тогда выполняется строгое неравенство:

$$F(d_j, e_j) < F(d_j, e_{j+1}).$$

Положить $d_{j+1} := e_{j+1}^{\top}$; $J := J \cup \{j+1\}$. Вычислить элементы $F_{j,j+1}$, $F_{j+1,j+1}$ матрицы значений $F[J,J]$. В соответствии с неравенством (30),

$$F_{jj} + F_{j+1,j+1} \geqslant 2F_{j,j+1}.$$

Тогда $F_{j+1,j+1} > F_{j,j+1}$ и $F_{j+1,j+1} > F_{j,j}$. Перейти к шагу 1, увеличив переменную цикла j на единицу.

14°. На каждой итерации описанного алгоритма значение целевой функции F строго возрастает. Все пары $\{d_j, e_j\}$, $j \in J$, есть пары базисных решений транспортной задачи, их конечное число. Стало быть, алгоритм конечен.

Если процесс остановился на паре $\{d_i, e_i\}$, то матрица $e_i[N,P]$ очевидно есть стационарная точка Q-задачи. Кроме того, это базисное решение. Базисная стационарная точка Q-задачи может быть одного из следующих двух видов [6]:

(a) локальный максимум в области допустимых значений;

(b) глобальный минимум на аффинной оболочке грани любой ненулевой размерности допустимого многогранника — в том случае, если этот минимум совпал с некоторой вершиной. Можно подобрать ограничения задачи таким образом, чтобы одна из вершин $e_i[N,P]$ совпала с глобальным минимумом на аффинной оболочке некоторой грани. Тогда стационарная точка $e_i[N,P]$ может не оказаться локально оптимальной. Такая ситуация, однако, встречается редко. Таким образом, алгоритм "почти всегда" дает локальный максимум F-задачи.

Из описания алгоритма следует, что существует стационарная пара базисных решений F-задачи. Нетрудно показать [5], что существует глобально оптимальная пара базисных решений. В наиболее употребительной постановке задачи кластерного анализа $\mu[N] = \mathbf{1}[N]$; $m[P]$ — целые числа при всех $p \in P$. В этом случае базисные решения целочисленны. Тогда базисное решение $e[N, P] = \varepsilon[N, P] = \sigma[N, P]$ представляет собой разбиение $\varphi : N \to P$ (см п. 1°).

Описанный алгоритм был реализован на языке ПЛ/1 и испытан при $|N| \sim 128$, $|K| = 2$ и различных значениях $|P| \in 2{:}8$. Оказалось, что число итераций $|J|$ обычно примерно равно числу кластеров $|P|$.

15°. **Заключение.** В представленной постановке задачи кластерного анализа можно использовать известные алгоритмы для нахождения глобального оптимума. Поскольку заранее известно, что область допустимых решений ограничена, то можно применить алгоритм, аналогичный [4].

Алгоритмы выпуклой максимизации, доведенные до программной реализации, могут работать лишь с относительно небольшим количеством переменных [12]. В реальных задачах кластеризации количество точек $|N|$ очень велико ($|N| > 100$), зато размерность пространства $|K|$ и число кластеров $|P|$, как правило, не превосходит 10. Ранг матрицы $R[N, N]$ не превосходит $|K|$. Этот факт можно использовать при решении задачи.

16°. Значительно более интересна задача, где массы кластеров $m[P]$ (3) не фиксированы. В этом случае дополнительная функция F (23), а также функция Q (31) принимают дробно-квадратичный вид. Задача глобальной оптимизации для дробно-квадратичной функции также рассматривалась в литературе [12]. Установлено, что, если квадратичная форма в числителе положительно определена, то:

(a) целевая функция квазивыпукла;

(b) существует базисное оптимальное решение;

(c) алгоритмы выпуклой максимизации, использующие ранжирование вершин [10], без существенных изменений можно перенести на рассматриваемую задачу [13].

Автор благодарит В. Н. Малозёмова за помощь в работе над статьей.

ЛИТЕРАТУРА

1. М. К. Гавурин, В. Н. Малоземов. *Экстремальные задачи с линейными ограничениями.* Ленинград, 1984. 176 с.

2. Б. Дюран, П. Оделл. *Кластерный анализ.* М., 1977. 128 с.

3. *Методы анализа данных: подход, основанный на методе динамических сгущений.* Кол. авт. под рук. Э. Дидэ. М., 1985. 360 с.

4. Е. Н. Сокирянская. *Некоторые алгоритмы билинейного программирования.* "Кибернетика", № 4, 1974.

5. M. Altman. *Bilinear Programming* "Бюллетень польской академии наук. Серия математических, астрономических и физических наук", 16, № 9, 1968.

6. C. A. Burdet. *Elements of a Theory in Non-convex Programming.* "Naval Research Logistic Quarterly", 24, № 1, 1977.

7. J. Czochralska. *The Method of Bilinear Programming for Non-convex Quadratic Programming.* "Zastosowania Mathematyki", 17, № 3, 1982.

8. G. Gallo, A. Ülkücü. *Bilinear Programming: an Exact Algorithm.* "Mathematical Programming", 12, № 3, 1977.

9. H. Konno. *Maximization of a Convex Quadratic Function under Linear Constraints.* "Mathematical Programming", 11, № 1, 1976.

10. K. G. Murty. *Solving the Fixed Charge Problem by Ranking the Extreme Points.* "Operations Research", 16, № 2, 1968.

11. P. M. Pardalos. *An Algorithm for a Class of Nonlinear Fractional Problems Using Ranking of the Vertices.* "BIT", 26, № 3, 1986.

12. P. M. Pardalos, J. B. Rosen. *Methods for Global Concave Minimization: a Bibliographic Survey.* "SIAM Review", 22, № 3, 1986.

13. S. Storøy. *Ranking of Vertices in the Linear Fractional Programming Problem.* "BIT", 23, № 3, 1983.

14. T. V. Thieu. *Relationships between Bilinear Programming and Concave minimization under Linear Constraints.* "Acta Mathematica Vietnam", 5, № 2, 1980.

15. H. Vaish, K. Shetty. *The Bilinear Programming Problem.* "Naval Research Logistic Quarterly", 23, № 2, 1976.