Математическая Теория Игр и её Приложения, т.11, в.4, с. 24–43

УДК 519.834
ББК 22.18

ДВУХФАКТОРНОЕ DEA МОДЕЛИРОВАНИЕ И КЛАСТЕРИЗАЦИЯ МНОЖЕСТВА ОДНОРОДНЫХ ФИРМ*

Владимир М. Буре[a]
Елена М. Парилина[a,b,c]
Ксения Ю. Староверова[a]
[a] Санкт-Петербургский государственный университет
199034, Санкт-Петербург, Университетская наб., 7-9
[b] Школа математики и статистики
Университет Циндао
Циндао 266071, Китай
[c] Институт прикладной математики провинции Шаньдун
Циндао 266071, Китай
e-mail: v.bure@spbu.ru, e.parilina@spbu.ru, st016234@student.spbu.ru

В работе представлена модель кластеризации однородных фирм по эффективности их деятельности на некотором временном интервале. Эффективность деятельности фирм проводится с использованием DEA (Data envelopment analysis) методологии, которая основана на решении оптимизационных задач и позволяет проводить сравнительный анализ фирм, учитывая множество факторов их деятельности. На втором этапе анализа полученные в результате DEA моделирования временные ряды оценок эффективности фирм используются для кластеризации фирм и нахождения устойчивых разбиений совокупности фирм.

* Работа второго автора поддержана грантом провинции Шаньдун (Shandong Province) “Double-Hundred Talent Plan” (No. WST2017009).

Ключевые слова: DEA моделирование, кластеризация временных рядов, анализ эффективности фирм.

Поступила в редакцию: 29.03.19 *После доработки*: 16.07.19 *Принята к публикации*: 30.07.19

1. Введение

В последнее время широкое применение получила методология DEA (Data envelopment analysis или анализ среды функционирования) моделирования, предложенная в работе [8]. Метод DEA позволяет найти оценки эффективности объектов, учитывая несколько показателей их деятельности, путем решения специальным образом сформулированных оптимизационных задач. Впоследствии было предложено несколько его модификаций, например, см. [6]. Главной задачей DEA моделирования является оценка технической эффективности работы фирмы среди совокупности достаточно однородных фирм и последующего упорядочения рассматриваемых фирм с точки зрения эффективности их работы. При этом оценивается прежде всего эффективность работы менеджмента фирмы или, другими словами, эффективность DMU (decision making unit), осуществляющей управление капиталом фирмы и трудовыми ресурсами. В настоящее время метод DEA используется повсеместно наряду со стандартными статистическими процедурами при анализе эффективности объектов, которые могут быть предприятиями, регионами, странами. Например, можно использовать методологию DEA при ранжировании грантозаявителей, учитывая их показатели деятельности в предыдущем периоде. Пример построения DEA модели при распределении ресурсов в разветвленной сети производства предложен в статье [13]. Метод оценки эффективности деятельности российских банков с использованием DEA методологии представлен в [1]. Проблема выбора эффективного портфеля, рассмотренная в [11], также использует анализ среды функционирования.

Нами предлагается метод нахождения устойчивых разбиений множества фирм в два этапа. На первом этапе проводится DEA моделирования деятельности фирм на некотором временном интервале. В результате первого этапе получается набор временных рядов оценок эффективностей фирм. На втором этапе предлагается использовать процедуру кластеризации временных рядов для нахождения разбиения множества фирм на непересекающиеся подмножества «близких» по эффективности фирм. Процедура кластеризации временных рядов изложена в [4]. Помимо решения задачи разбиения множества фирм на кластеры, мы предлагаем применить теоретико-игровой метод нахождения устойчивого разбиения, изложенный в [14], а впоследствии адаптированный для нахождения устойчивого множества кластеров в работе [2]. Под устойчивостью разбиения множества фирм на кластеры понимается невозможность улучшить целевую функцию, специальным образом зависящую от межкластерного и внутрикластерного расстояния, путем переноса какого-либо объекта из одного кластера в другой или выделением его в отдельный кластер. В последнее время теоретико-игровые модели стали применяться в кластеризации данных, например, при разбиении научных сообществ [3], нахождении разбиений в сети заданной конфигурации [7]. Элементы теории игр, в том числе, теории нечетких кооперативных игр, используются и при DEA моделировании [12,13]. В упомянутых работах проводится анализ «эффективностей» городов в Иране и эффективностей работы медицинских учреждений.

Статья имеет следующую структуру. В разделе 2 приведены теоретические основы DEA моделирования. Метод кластеризации полученных в результате DEA анализа временных рядов, основанный на теоретико-игровом подходе, представлен в разделе 3. Пример в разделе 4 иллюстрирует предложенную методологию анализа эффективности фирм. Описание результатов работы проводится в Заключении.

2. DEA моделирование

Как известно, двумя основными факторами производства являются труд и капитал. В экономических исследованиях часто используются двухфакторные производственные функции Кобба-Дугласа [10], имеющие следующий вид

$$Y = AK^{\alpha}L^{\beta}, \tag{2.1}$$

где Y – произведенная стоимость (стоимость выпущенной продукции), фактор K – капитал (или стоимость основных фондов), фактор L – труд (в денежном выражении может характеризовать размер выплат всем участникам производственного процесса), α – константа (коэффициент эластичности производства по капиталу K), β – коэффициент эластичности производства по труду L, A – коэффициент, который часто связывают с уровнем технологий.

Предположим, что имеется n предприятий (фирм), выпускающих однотипную продукцию или оказывающих однотипные услуги (например, отели). Деятельность каждой из фирм за некоторый фиксированный период времени (например, квартал) будем характеризовать набором параметров Y_i, K_i, L_i, $(i = 1, \ldots, n)$, имеющих тот же смысл, что и в производственной функции Кобба-Дугласа.

Для оценки технической эффективности фирм и последующего составления рейтинга применим подход на основе DEA моделирования. Будем характеризовать эффективность работы фирмы за рассматриваемый период времени величиной

$$T_i(\delta_0, \delta_1, \delta_2) = \frac{\delta_0 Y_i}{\delta_1 K_i + \delta_2 L_i}, \qquad i = 1, \ldots, n. \tag{2.2}$$

Введем множество D допустимых значений весовых коэффициентов $(\delta_0, \delta_1, \delta_2)$:

$$D = \left\{(\delta_0, \delta_1, \delta_2) \,|\, \delta_0 \geqslant 0, \delta_1 \geqslant 0, \delta_2 \geqslant 0, \quad T_i(\delta_0, \delta_1, \delta_2) \leqslant 1,\, i = 1, \ldots, n\right\}. \tag{2.3}$$

Рассмотрим n оптимизационных задач:

$$\max_{(\delta_0, \delta_1, \delta_2) \in D} T_i(\delta_0, \delta_1, \delta_2), \quad i = 1, \ldots, n. \tag{2.4}$$

Предположим, что набор весовых коэффициентов $(\delta_{0i}^*, \delta_{1i}^*, \delta_{2i}^*)$ является решением сформулированной задачи (2.4), то есть

$$T_i^* = T_i(\delta_{0i}^*, \delta_{1i}^*, \delta_{2i}^*) = \max_{(\delta_0, \delta_1, \delta_2) \in D} T_i(\delta_0, \delta_1, \delta_2). \tag{2.5}$$

Назовем величину T_i^* показателем (коэффициентом) технической эффективности фирмы $i = 1, \ldots, n$. Очевидно, что коэффициенты $T_i(\delta_{0i}^*, \delta_{1i}^*, \delta_{2i}^*) \leqslant 1$, $i = 1, \ldots, n$. Коэффициенты технической эффективности очевидным образом упорядочиваются в виде вариационного ряда[1] значений коэффициентов технической эффективности всех

---

[1]Упорядочим коэффициенты T_i^* в порядке невозрастания и присвоим им новые номера $T_{(i)}^*$ в круглых скобках, которые будут использоваться в дальнейшем для упорядоченных выборок (вариационных рядов).

n фирм:

$$T^*_{(1)} \geqslant T^*_{(2)} \geqslant \ldots \geqslant T^*_{(k)} \geqslant T^*_{(k+1)} \geqslant \ldots \geqslant T^*_{(n)}. \tag{2.6}$$

Ряд (2.6) представляет собой рейтинг рассматриваемой совокупности однородных фирм для данного периода времени. Предположим, что первые k коэффициентов вариационного ряда равны единице, то есть

$$T^*_{(1)} = T^*_{(2)} = \ldots = T^*_{(k)} = 1,$$

тогда в соответствии с методологией DEA моделирования соответствующие этим коэффициентам фирмы рассматриваются как фирмы с максимальной технической эффективностью (оптимально функционирующие) в данной совокупности фирм на рассматриваемом промежутке времени.

Преобразуем выражения для коэффициентов $T_i(\delta_0, \delta_1, \delta_2)$, $i = 1, \ldots, n$ следующим образом:

$$T_i(\delta_0, \delta_1, \delta_2) = \frac{\delta_0 Y_i}{\delta_1 K_i + \delta_2 L_i} = \frac{1}{\theta_1 x_{1i} + \theta_2 x_{2i}} = \frac{1}{S_i(\theta_1, \theta_2)},$$

где

$$\theta_1 = \frac{\delta_1}{\delta_0}, \quad \theta_2 = \frac{\delta_2}{\delta_0},$$
$$x_{1i} = \frac{K_i}{Y_i}, \quad x_{2i} = \frac{L_i}{Y_i},$$
$$S_i(\theta_1, \theta_2) = \theta_1 x_{1i} + \theta_2 x_{2i}.$$

Введем в рассмотрение множество

$$\tilde{D} = \big\{(\theta_1, \theta_2) \,|\, \theta_1 \geqslant 0, \theta_2 \geqslant 0,\ S_i(\theta_1, \theta_2) \geqslant 1,\ i = 1, \ldots, n\big\}.$$

Рассмотрим оптимизационные задачи следующего вида:

$$\min_{(\theta_1, \theta_2) \in \tilde{D}} S_i(\theta_1, \theta_2), \quad i = 1, \ldots, n. \tag{2.7}$$

Пусть набор $(\theta^*_{1i}, \theta^*_{2i})$ является решением i-ой оптимизационной задачи (2.7):

$$S^*_i = S_i(\theta^*_{1i}, \theta^*_{2i}) = \min_{(\theta_1, \theta_2) \in \tilde{D}} S_i(\theta_1, \theta_2), \quad i = 1, \ldots, n. \tag{2.8}$$

Для значений n оптимизационных задач S_i^*, $i = 1, \ldots, n$ можно построить вариационный ряд

$$S_{(1)}^* \leqslant S_{(2)}^* \leqslant \ldots \leqslant S_{(k)}^* \leqslant S_{(k+1)}^* \leqslant \ldots \leqslant S_{(n)}^*, \tag{2.9}$$

который эквивалентен рейтингу (2.6). В системе декартовых координат (θ_1, θ_2) на плоскости рассмотрим n прямых в первом квадранте, $(\theta_1 \geqslant 0, \theta_2 \geqslant 0)$, определяемых уравнениями

$$S_i(\theta_1, \theta_2) = 1, \qquad i = 1, \ldots, n. \tag{2.10}$$

Считая, что $x_{2i} > 0$, уравнения (2.10) можно разрешить относительно θ_2, в результате получаем линейные функции вида:

$$\theta_2 = -b_i\theta_1 + a_i, \qquad i = 1, \ldots, n, \tag{2.11}$$

где

$$b_i = \frac{x_{1i}}{x_{2i}}, \qquad a_i = \frac{1}{x_{2i}}.$$

Рассмотрим кусочно-линейную огибающую этих прямых

$$\hat{\theta}_2(\theta_1) = \max_{i=1,\ldots,n} (-b_i\theta_1 + a_i), \tag{2.12}$$

где $\theta_1 \in \left[0, \max\limits_{i=1,\ldots,n} \left\{\frac{1}{x_{1i}}\right\}\right]$.

Теорема 2.1. *Справедливы следующие утверждения:*

1. *Огибающая* $\hat{\theta}_2(\theta_1)$ *на множестве* $\left[0, \max\limits_{i=1,\ldots,n} \left\{\frac{1}{x_{1i}}\right\}\right]$ *является вогнутой функцией аргумента* θ_1.

2. *Множество* $\tilde{D} = \left\{(\theta_1, \theta_2) \,|\, \theta_1 \geqslant 0, \theta_2 \geqslant 0,\ S_i(\theta_1, \theta_2) \geqslant 1,\ i = 1, \ldots, n\right\}$ *совпадает с множеством*

$$\left\{(\theta_1, \theta_2) \,|\, \theta_1 \in \left[0, \max_{i=1,\ldots,n} \left\{\frac{1}{x_{1i}}\right\}\right], \theta_2 \geqslant \hat{\theta}_2(\theta_1)\right\}$$

3. *Для оптимизационных задач* (2.4) *и* (2.7) *выполняется равенство*

$$T_i^* = \frac{1}{S_i^*}, \quad i = 1, \ldots, n.$$

4. *Решения задач* (2.4) *и* (2.7) *существуют.*

5. *Выполняется равенство* $T^*_{(1)} = S^*_{(1)} = 1$.

Доказательство. Утверждение 1 теоремы очевидным образом следует из определения функции $\hat{\theta}_2(\theta_1)$ по формуле (2.12). Второе утверждение теоремы следует из построения вспомогательной оптимизационной задачи вида (2.7). Условие $S_i(\theta_1, \theta_2) \geqslant 1$ эквивалентно условию $\theta_2 \geqslant \hat{\theta}_2(\theta_1)$. Условия $\theta_1 \geqslant 0$, $\theta_2 \geqslant 0$ эквивалентны условию $\theta_1 \in \left[0, \max\limits_{i=1,\ldots,n} \left\{\frac{1}{x_{1i}}\right\}\right]$. По построению эквивалентной оптимизационной задачи (2.7), описанной выше, очевидно, что решения задач (2.4) и (2.7) удовлетворяют соотношению $T^*_i = \dfrac{1}{S^*_i}$ для всех $i = 1, \ldots, n$. Поскольку область D – непустое множество, то решение задачи (2.4) существует, из чего следует существование решения задачи (2.7) в силу их эквивалентности. Так как сформулированные оптимизационные задачи являются задачами линейного программирования (2.4) с ограничениями в виде области D, где $T_i \leqslant 1$, то можно утверждать, то существует хотя бы одно i, для которого $T^*_i = 1$. Следовательно, в вариационном ряду $T^*_{(1)} = 1$. В силу эквивалентности оптимизационных задач (2.4) и (2.7), значению $T^*_{(1)} = 1$ соответствует значение $S^*_{(1)} = 1$ в вариационном ряду (2.9). □

Следствие 2.1. *Для значений индекса* i*, таких что при каком-либо* $\theta_1 \in \left[0, \max\limits_{i=1,\ldots,n} \left\{\frac{1}{x_{1i}}\right\}\right]$ *и* $\hat{\theta}_2(\theta_1) = -b_i\theta_1 + a_i$*, где* $b_i = \frac{x_{1i}}{x_{2i}}$, $a_i = \frac{1}{x_{2i}}$*, выполняется равенство* $T^*_i = S^*_i = 1$*, фирма* i *в данной совокупности фирм обладает максимальной эффективностью.*

Рассмотрение, проведенное выше, относится к некоторому фиксированному промежутку времени (например, к некоторому кварталу определенного года). Разумеется, целесообразно исследование динамики рейтингов фирм [3], другими словами, для последовательных промежутков времени $m = 1, \ldots, M$, аналогично предыдущему строятся оценки технической эффективности каждой из фирм среди рассматриваемой совокупности фирм для последовательных временных интервалов. Таким образом, для каждой фирмы $i = 1, \ldots, n$ строится последовательность коэффициентов T^*_{im}, $m = 1, \ldots, M$, характеризующих техническую эффективность фирмы i среди совокупности рассматриваемых фирм для последовательных промежутков времени (одинаковой длины) $m = 1, \ldots, M$. Вместо рассмотрения коэффициентов T^*_{im}, $m = 1, \ldots, M$, можно исследовать эквивалентные им коэффициенты S^*_{im}, $m = 1, \ldots, M$.

Следовательно, каждая фирма i порождает временной ряд коэффициентов, характеризующих техническую эффективность фирмы. Особенность этого временного ряда в том, что имеется идеальная точка, то есть ряд, состоящий из единиц, так как максимальное (наилучшее) значение коэффициента равно единице. Кластеризация фирм и поиск устойчивых коалиционных разбиений полностью основывается на кластеризации временных рядов коэффициентов технической эффективности.

3. Теоретико-игровой подход в кластерном анализе

Кластеризация – это задача разбиения объектов на группы таким образом, чтобы внутри каждой группы оказались похожие (относительно заранее заданного функционала) объекты, а объекты разных групп должны быть как можно более отличны. Известно, что результаты кластеризации зависят от многих факторов: выбора метрики или меры близости, подхода кластеризации, методов определения количества кластеров. В рамках данной работы предлагается использовать теоретико-игровой подход [2], так как он позволяет найти устойчивую кластерную структуру.

3.1. Игры с коалиционной структурой

Метод кластеризации, предложенный в [2], основанный на теоретико-игровом подходе, поэтому напомним сначала основные определения и теоремы, используемые далее в работе. Рассмотрим игру (N, v, π) со множеством игроков N, характеристической функцией v и коалиционной структурой π.

Определение 3.1. *Характеристической функцией в игре с множеством игроков N называется функция $v : 2^N \to \mathbb{R}^1$, определенная на множестве подмножеств N, где $v(S)$, $S \subset N$, показывает «силу» коалиции S. При этом, $v(\varnothing) = 0$.*

Определение 3.2. *Коалиционная структура* π *– это разбиение множества всех игроков* N *следующим образом:* $\pi = \{B_1, \ldots, B_m\}$, *причем*

- $B_1 \cup B_2 \cup \cdots \cup B_m = N$,
- $B_i \cap B_j = \varnothing, \forall i, j = 1, \ldots, m,\ i \neq j$.

Определение 3.3. *Вектор* $\psi^\pi = (\psi_1^\pi, \ldots, \psi_n^\pi) \in \mathbb{R}^n$ *называется распределением выигрыша в игре* (N, v, π), *если выполняется условие коллективной рациональности:*

$$\sum_{i \in B_j} \psi_i^\pi = v(B_j), \quad \forall B_j \in \pi.$$

Определение 3.4. *Распределение выигрыша* $\psi^\pi \in \mathbb{R}^n$ *называется дележом в игре* (N, v, π), *если выполняется условие индивидуальной рациональности:*

$$\psi_i^\pi \geqslant v(\{i\}), \quad \forall i \in N.$$

В кооперативной теории игр существует ряд процедур для распределения выигрыша ψ^π, в этой работе будут рассмотрены вектор Ауманна–Дрезе

$$\phi_k^\pi = \sum_{\substack{S \subseteq B(k) \\ S \ni k}} \frac{(|B(k)| - |S|)!(|S| - 1)!}{|B(k)|!} \Big[v(S) - v(S \setminus \{k\}) \Big]$$

и ES-вектор (equal surplus division value)

$$\xi_k^\pi = v(\{k\}) + \frac{\Big(v(B(k)) - \sum\limits_{i \in B(k)} v(\{i\}) \Big)}{|B(k)|},$$

где $B(k) \in \pi$ – коалиция, содержащая игрока $k \in N$.

Практический интерес представляет поиск такой коалиционной структуры, от которой не выгодно отклоняться ни одному игроку. Наличие такой коалиционной структуры свидетельствует о существовании четкой кластеризации. Определение устойчивой коалиционной структуры может быть сформулировано различными способами, в этой работе использовано определение из работы [15], так как оно соответствует концепции равновесия Нэша.

Определение 3.5. *Коалиционную структуру π будем называть устойчивой относительно выбранного одноточечного принципа оптимальности, если для любого игрока $k \in N$, любой коалиции $B(k) \in \pi$ и коалиционной структуры π':*

$$\psi_k^{\pi} \geqslant \psi_k^{\pi'},$$

где разбиение π' отличается от π тем, что игрок k переходит из коалиции $B(k)$ в B, т. е. $\pi' = \{B(k) \setminus \{k\}, B \cup \{k\}, \pi_{-B(k) \cup B}\}$, причем $B \in \pi_{-B(k)} \cup \varnothing$ и π_{-B_k}, такое что $\pi_{-B_k} = \pi \setminus B_k$.

Представим, что коалиционная структура π устойчива и игрок $k \in B(k)$ имеет следующие стратегии:

1. Остаться в коалиции $B(k)$.
2. Сформировать новую коалицию $\{k\}$.
3. Присоединиться к другой существующей коалиции в π.

Тогда игрок k не может получить больший выигрыш, придерживаясь стратегий 2 и 3, чем он получит при стратегии 1, при условии, что остальные игроки не отклоняются. Из этих принципов вытекает следующая теорема, доказательство которой приведено в [4].

Теорема 3.1. *Для того, чтобы коалиционная структура π была устойчивой относительно выбранного принципа оптимальности необходимо, чтобы распределение выигрыша ψ^{π}, вычисленное согласно этому принципу оптимальности для коалиционной структуры π, являлось дележом.*

3.2. Кластеризация как задача теории игр

Пусть $N = \{1, 2, \dots, n\}$ – множество объектов таких, что для любых двух объектов определена функция расстояния $d(i,j) : N \times N \mapsto [0,1]$ и функция сходства $f(i,j) : N \times N \mapsto [0,1]$. В задаче кластеризации коалицию $S \subset N$ будем называть кластером. Причем для этих функций выполнено свойство симметричности, т. е. $d(i,j) = d(j,i)$ и $f(i,j) = f(j,i)$ при всех $i, j \in N$ и с возрастанием одной функции вторая не возрастает:

$$d(i,j) \geqslant d(k,l) \Longleftrightarrow f(i,j) \leqslant f(k,l) \text{ для всех } i,j,k,l \in N.$$

Под коалицией S понимается некоторое непустое подмножество множества объектов N, которые объединяются в кластер. Рассмотрим характеристическую функцию игры (N, v, π) в виде

$$v(S) = \sum_{\substack{i,j \in S \\ i \neq j}} f(i,k) + 2 \sum_{\substack{i \in S \\ k \in N \setminus S}} d(i,j). \tag{3.1}$$

Значение характеристической функции $v(S)$ велико для кластеров S, которые содержат такие объекты, которые имеют большое значение функции сходства $f(i,j)$ для объектов, входящих в кластер $i, j \in S$, и в то же время большое расстояние $d(i,k)$ до объектов других кластеров $k \in N \setminus S$.

Характеристическая функция $v(\cdot)$, определяемая согласно (3.1), не является монотонной, т. е. неравенство $v(S) \leqslant v(T)$ не выполняется для любых двух коалиций $S \subset N$ и $T \subset N$, таких что $S \subseteq T$. Характеристическая функция $v(\cdot)$, определяемая согласно (3.1), не является супераддитивной, т. е. неравенство $v(S \cup T) \geqslant v(S) + v(T)$ не выполняется для любых двух непересекающихся непустых коалиций $S \subset N$ и $T \subset N$. Также значение характеристической функции объединения двух кластеров $v(S \cup T)$ будет больше, чем сумма значений характеристических функций на кластерах $v(S) + v(T)$, только при условии, что сходство объектов хотя бы в 2 раза больше, чем расстояние между ними $\sum_{\substack{i \in S \\ j \in T}} f(i,j) \geqslant 2 \sum_{\substack{i \in S \\ j \in T}} d(i,j)$. Эти свойства построенной характеристической функции вида (3.1) доказаны в [2].

Отсутствие свойства супераддитивности говорит о том, что в общем случае игрокам не выгодно объединяться в гранд-коалицию, что естественно для задачи кластеризации, и именно поэтому имеет смысл рассматривать игры с коалиционной структурой π, коалиционную структуру можно интерпретировать как возможный вариант кластеризации. Таким образом, задача кластеризации сводится к поиску устойчивой коалиционной структуры.

3.3. Существование устойчивых коалиционных структур

Утверждение 3.1. *В игре (N, v, π) с коалиционной структурой π и характеристической функцией v вида (3.1) компоненты вектора*

Ауманна–Дрезе ϕ^π определяются формулой:

$$\phi_k^\pi = \sum_{\substack{S\subseteq B(k)\\ S\ni k}} \frac{2(b-s)!(s-1)!}{b!}\Big[\sum_{i\in S\setminus\{k\}}\big(f(i,k)-d(i,k)\big)+\sum_{i\in N\setminus S} d(i,k)\Big],$$

где $|B(k)| = b$ и $|S| = s$.

Следствие 3.1. $\phi_k^\pi \geqslant 0$ *тогда и только тогда, когда выполнено следующее неравенство*

$$\sum_{\substack{S\subseteq B(k)\\ S\ni k}} \frac{2(b-s)!(s-1)!}{b!}\Big[\sum_{i\in S\setminus\{k\}}\big(f(i,k)-d(i,k)\big)+\sum_{i\in N\setminus S} d(i,k)\Big] \geqslant 0.$$

Можно заметить, что в тривиальном случае, когда для некоторого игрока $k \in N$ и всех $S \subseteq B(k)$, содержащих k, выполнено неравенство

$$f(i,k) \geqslant d(i,k), \forall i \in S\setminus\{k\},$$

компоненты вектора Ауманна–Дрезе ϕ_k^π будут неотрицательны.

Утверждение 3.2. *Коалиционная структура π в игре (N, v, π) устойчива относительно вектора Ауманна–Дрезе, если*

$$\forall k \in N, \forall B \in \pi_{-B(k)} \cup \varnothing \implies$$

$$\sum_{\substack{S\subseteq B(k)\\ S\ni k}} \frac{2(b-s)!(s-1)!}{b!}\Big[\sum_{i\in S\setminus\{k\}}\big(f(i,k)-d(i,k)\big)+\sum_{i\in N\setminus S} d(i,k)\Big] \geqslant$$

$$\sum_{\substack{S'\subseteq B\cup\{k\}\\ S'\ni k}} \frac{2(b'-s')!(s'-1)!}{b'!}\Big[\sum_{i\in S'\setminus\{k\}}\big(f(i,k)-d(i,k)\big)+\sum_{i\in N\setminus S'} d(i,k)\Big],$$

здесь $|B(k)| = b, |S| = s$, $|B| = b', |S'| = s'$ и $B \cap B(k) = \varnothing$, $B \in \pi \cup \varnothing$, $B(k) \in \pi$.

Замечание 3.1. В игре (N, v, π) не всегда существует устойчивая относительно вектора Ауманна–Дрезе коалиционная структура.

Утверждение 3.3. *В игре (N, v, π) с коалиционной структурой π и характеристической функцией v вида (3.1) компоненты ES-вектора*

ξ^π заданы формулой:

$$\xi_k^\pi = 2\sum_{j\in N\setminus\{k\}} d(k,j) + \frac{1}{b}\Big(\sum_{\substack{i\in B(k)\\ j\in B(k)\\ i\neq j}} \big(f(i,j) - 2d(i,j)\big)\Big),$$

где $|B(k)| = b$.

Следствие 3.2. *$\xi_k^\pi \geqslant 0$ тогда и только тогда, когда выполнено следующее неравенство*

$$2\sum_{j\in N\setminus\{k\}} d(k,j) \geqslant \frac{1}{b}\Big(\sum_{\substack{i\in B(k)\\ j\in B(k)\\ i\neq j}} \big(2d(i,j) - f(i,j)\big)\Big).$$

Можно заметить, что в случае объединения в кластер сильно схожих объектов выполняется условие неотрицательности компонент вектора (3.2). Под сильно схожими объектами $i, j \in N$ здесь понимаются такие, что выполняется $f(i,j) \geqslant 2d(i,j)$.

Утверждение 3.4. *Коалиционная структура π в игре (N, v, π) устойчива относительно ES-вектора, если*

$$\forall B \in \pi, \forall B' \in \pi_{-B} \cup \varnothing \implies$$

$$\frac{1}{b}\Big[\sum_{\substack{i\in B\\ j\in B\\ i\neq j}} \big(f(i,j) - 2d(i,j)\big)\Big] \geqslant \frac{1}{b'}\Big[\sum_{\substack{i\in B'\\ j\in B'\\ i\neq j}} \big(f(i,j) - 2d(i,j)\big)\Big],$$

$|B| = b$, $|B'| = b'$ и $B \cap B' = \varnothing$, $B' \in \pi \cup \varnothing$, $B \in \pi$.

Следствие 3.3. *Из предыдущего утверждения и теоремы 3.1 следует, что необходимым условием устойчивости коалиционной структуры π в игре (N, v, π) является выполнение следующего неравенства:*

$$\frac{1}{b}\Big[\sum_{\substack{i\in B\\ j\in B\\ i\neq j}} \big(f(i,j) - 2d(i,j)\big)\Big] \geqslant 0$$

для всех B в коалиционной структуре π.

Замечание 3.2. В игре (N, v, π) не всегда существует устойчивая относительно вектора ES-вектора структура.

Все утверждения раздела 3.3 доказаны в работе [2].

4. Пример

В этом разделе представлен результат работы предложенного метода по нахождению кластеров фирм-пекарен с учетом эффективности их работы на протяжении 20 лет. Пусть каждая пекарня имеет номер i от 0 до 9 и заданы Y_i, K_i, L_i – произведенная стоимость, стоимость основных фондов и размер выплат членам трудового коллектива соответственно как основные характеристики их деятельности за последние 20 лет (1998–2018). Эти характеристики представлены на рис. 1–3.

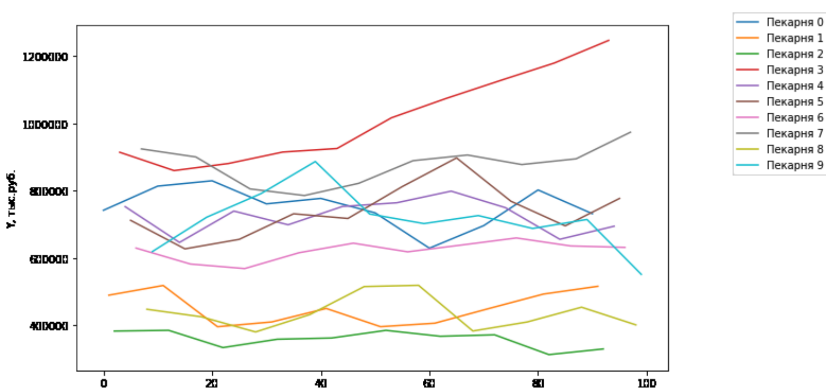

Рисунок 1. Характеристики эффективности пекарен: стоимость выпущенной продукции

Решив оптимизационные задачи вида (2.8) для каждого года, получим набор временных рядов, каждый ряд соответствует некоторой пекарне и содержит величины, обратные коэффициентам технической эффективности (см. полученные временные ряды на рис.4).

Для того чтобы найти пекарни с похожей эффективностью, мы будем решать задачу кластеризации временных рядов $S^*_{(i,t)}$, $i = \{0, 1, \ldots, 9\}$, $t = \{1998, 1999, \ldots, 2018\}$. В качестве функции расстояния выбрана мера различия d, основанная на временной корреляции и

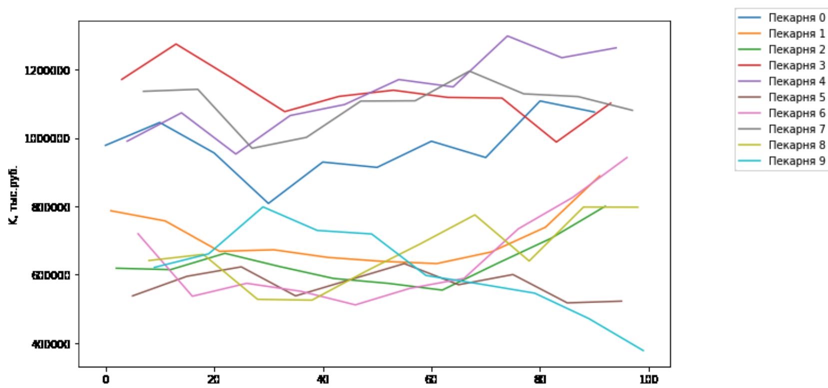

Рисунок 2. Характеристики эффективности пекарен: капитал

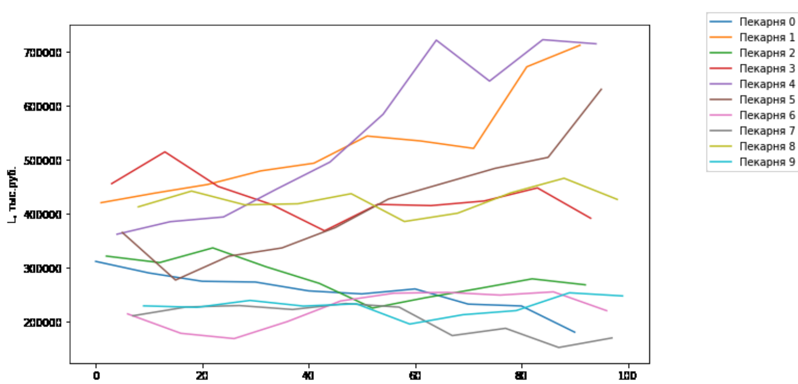

Рисунок 3. Характеристики эффективности пекарен: оплата труда

евклидовом расстоянии (CORT) с параметром k, равным единице [8]:

$$\mathrm{c}(x,y) = \frac{\sum\limits_{t=1}^{n-1}\big(x(t+1)-x(t)\big)\big(y(t+1)-y(t)\big)}{\sqrt{\sum\limits_{t=1}^{n-1}\big(x(t+1)-x(t)\big)^2}\sqrt{\sum\limits_{t=1}^{n-1}\big(y(t+1)-y(t)\big)^2}},$$

$$d(x,y) = \frac{2}{1+\exp\left(kc(x,y)\right)}||x-y||_2.$$

т.к. при поиске расстояния она учитывает динамику временных рядов. Теоретико-игровая процедура кластеризации показала, что для

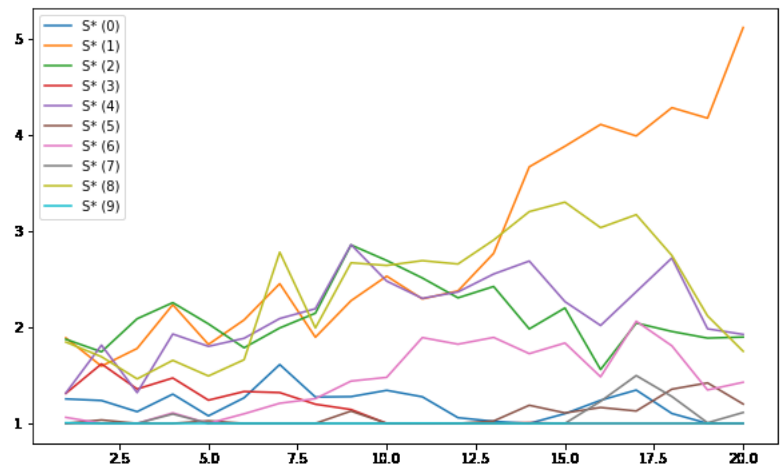

Рисунок 4. Решения оптимизационных задач (2.8) для пекарен

такой задачи существует устойчивое разбиение относительно ES-вектора и вектора Ауманна–Дрезе:

$$K = \{\{1\}, \{2, 4, 8\}, \{0, 3, 5, 6, 7, 9\}\}.$$

При этом ES-вектор для коалиционной структуры K выглядит следующим образом:

$$\xi^K = \{8.36, 14.43, 8.87, 9.57, 8.27, 9.56, 8.77, 8.95, 9.53, 9.95\}.$$

Вектор Ауманна–Дрезе для коалиционной структуры K равен

$$\phi_K = \{8.75, 14.43, 8.81, 9.52, 8.49, 9.67, 7.40, 9.57, 9.36, 10.22\}.$$

На рис. 5 временные ряды с величинами обратными коэффициентам технической эффективности разбиты на кластеры (каждый кластер имеет свой цвет и тип линии), можно заметить, что пекарни $\{0, 3, 5, 6, 7, 9\}$ образуют кластер наиболее эффективных пекарен, $\{2, 4, 8\}$ показывают среднюю эффективность и пекарня 1 образует отдельный кластер, так как в последние 7 лет демонстрирует ухудшение в работе.

Результаты DEA моделирования для пекарни 1 подтверждают первоначальную гипотезу о неэффективности работы пекарни 1. Если посмотреть на первоначальные характеристики работы пекарни 1

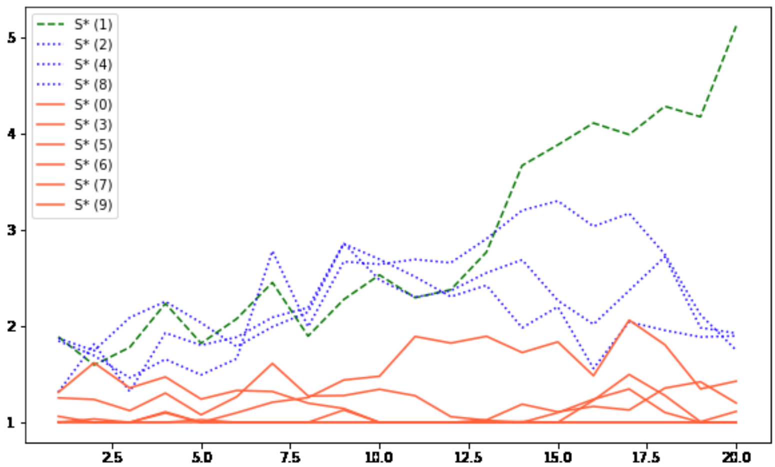

Рисунок 5. Разбиение пекарен на кластеры согласно эффективности их работы

в табл. 4, стоимость выпущенной продукции Y_1 убывает, в то время как величины L_1 и K_1 растут.

Таблица 1. Показатели эффективности работы пекарни 1 (нечетные года), тыс.руб

t	1999	2001	2003	2005	2007	2009	2011	2013	2015	2017
K_1	757	672	638	666	889	680	756	726	891	1283
L_1	437	479	543	521	712	634	658	819	893	906
Y_1	517	409	395	449	515	414	361	312	331	340

5. Заключение

В работе описана методология анализа деятельности предприятий – DEA моделирование. Такой подход особенно удобен для сравнения эффективности работы «однородных» объектов, так как методология нечувствительна к различиям в масштабе числовых характеристик. В связи с тем, что в финансовом анализе обычно интерес представляет динамика показателей, в статье предложен двухэтапный алгоритм: на первом этапе DEA моделирование позволяет получить временные ряды для каждого предприятия, характеризующие их эффективность, на втором – объекты разбиваются на группы при помощи методов кластеризации. Причем для многих задач возможно получить разбиение, которое будет устойчивым относительно выбранного одноточечного принципа оптимальности, отсутствие устойчивого разбиения говорит о несуществовании четкой кластеризации (в рамках определения 3.5). Таким образом, получая на входе характеристики функционирования предприятий, на выходе алгоритм выдает разбиение этих предприятий на группы объектов, схожих по динамике эффективности. Для наглядности в разделе 4 приведен пример анализа деятельности пекарен за последние 20 лет.

СПИСОК ЛИТЕРАТУРЫ

1. Бондаренко Ю. В., Глазьева Ю. В., Бутырина Н. А. *Алгоритм количественной оценки эффективности деятельности российских банков на основе подхода DEA* // Управление строительством. 2018. № 1 (10). С. 109–114.

2. Буре В. М., Староверова К. Ю. *Применение теории кооперативных игр с коалиционной структурой для кластеризации данных* // Математическая теория игр и ее приложения. 2018. Т. 10. № 1. С. 23–39.

3. Ермолин Н. А., Мазалов В. В., Печников А. А. *Теоретико-игровые методы нахождения сообществ в академическом Вебе* // Труды СПИИРАН. 2017. № 55. С. 237–254.

4. Парилина Е.М., Седаков А.А. *Устойчивость коалиционных структур одной модели банковской кооперации* // Математическая Теория Игр и ее Приложения. 2012. Т. 4, вып. 4. С. 45–62.

5. Староверова К. Ю., Буре В. М. *Мера различия временных рядов, основанная на их характеристиках* // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2017. Т. 13. № 1. С. 51–60.

6. Andersen P., Petersen N. C. *A Procedure for Ranking Efficient Units in Data Envelopment Analysis* // Management Science. 1993. Vol. 39. No. 10. P. 1261–1264.

7. Avrachenkov K. E., Kondratev A. Y, Mazalov V. V., Rubanov D. G. *Network partitioning algorithms as cooperative games* // Computational Social Networks. 2018. Vol. 5. No. 11. P. 1-28.

8. Charnes A., Cooper W. W., Rhodes E. *Measuring efficiency of decision making units* // European Journal of Operational Research. 1978. Vol. 2(6). P. 429–444.

9. Chouakria A.D., Nagabhushan P.N. *Adaptive dissimilarity index for measuring time series proximity* // Advances in Data Analysis and Classification. 2007. Vol. 1(1). Pp. 5–21.

10. Cobb C. W., Douglas P. H. *A Theory of Production* // American Economic Review. 1928. Vol. 18. No. 1. P. 139–165.

11. Fiala F. *Project portfolio designing using data envelopment analysis and De Novo optimisation* // Central European Journal of Operations Research. 2018. Vol. 26. P. 847–859.

12. Omrani H., Fahimi P., Mahmood A. *A data envelopment analysis game theory approach for constructing composite indicator: An application to find out development degree of cities in West Azarbaijan province of Iran* // Socio-Economic Planning Sciences (article in press)

13. Omrani H., Shafaat K., Emrouznejad A. *An integrated fuzzy clustering cooperative game data envelopment analysis model with application in hospital efficiency* // Expert Systems With Applications. 2018. Vol. 114. P. 615–628.

14. Shao Y., Bi G., Yang F., Xia Q. *Resource allocation for branch network system with considering heterogeneity based on DEA method* // Central European Journal of Operations Research. 2018. Vol. 26. P. 1005–1025.

15. Sedakov A., Parilina E., Volobuev Yu. & Klimuk D. *Existence of Stable Coalition Structures in Three-person Games* // Contributions to Game Theory and Management. 2013. Vol. 6. P. 407–422.

TWO-FACTOR DEA MODELING AND CLUSTARIZATION OF HOMOGENEOUS FIRMS

Vladimir M. Bure, Saint Petersburg State University, Dr.Sc., professor (v.bure@spbu.ru),
Elena M. Parilina, Saint Petersburg State University, School of Mathematics and Statistics, Qingdao University, Institute of Applied Mathematics of Shandong, Ph.D., associate professor (e.parilina@spbu.ru),
Kseniya Yu. Staroverova, Saint Petersburg State University, postgraduate student (st016234@student.spbu.ru).

Abstract: The paper presents a clasterization model for homogeneous firms using effectiveness of their activities over a certain time period. The firm efficiency is investigated by the DEA (Data envelopment analysis) methodology, which is based on solving optimization problems and allows to compare firms taking into account many factors of their activities. At the second step of the analysis, the time series of estimators of firm effectiveness obtained as a result of DEA modeling are used to cluster firms and find stable partitions of the set of firms.

Keywords: DEA modeling, time series clustarization, analysis of firm efficiencies.