Математическая Теория Игр и её Приложения, т.10, в.1, с. 23–39

УДК 519.834
ББК 22.18

ПРИМЕНЕНИЕ ТЕОРИИ КООПЕРАТИВНЫХ ИГР С КОАЛИЦИОННОЙ СТРУКТУРОЙ ДЛЯ КЛАСТЕРИЗАЦИИ ДАННЫХ

Владимир М. Буре
Ксения Ю. Староверова
Санкт-Петербургский государственный университет
199034, Санкт-Петербург, Университетская наб., 7–9
e-mail: v.bure@spbu.ru, st016234@student.spbu.ru

В работе рассматривается кооперативная игра, где игроками являются некоторые объекты, для которых можно численно измерить расстояние или сходство между ними. В такой игре можно определить характеристическую функцию таким образом, что ее значение будет велико для коалиций, содержащих наиболее близких (схожих) игроков и не содержащих совсем далеких (не схожих) от остальных игроков коалиции. Для такой функции нарушено свойство супераддитивности, поэтому целесообразно рассматривать модель с коалиционной структурой. Таким образом, мы можем рассматривать такую игру как процедуру кластеризации объектов (игроков). Полученные условия существования устойчивой коалиционной структуры позволяют определить наличие четкой кластеризации.

Ключевые слова: коалиция, устойчивая коалиционная структура, кластеризация, вектор Шепли, вектор Ауманна-Дрезе, ES-вектор.

©2018 В.М. Буре, К.Ю. Староверова

1. Введение

Кластеризация является важной задачей анализа данных. Предполагается, что нужно разбить объекты на группы таким образом, чтобы внутри каждой группы оказались похожие (в каком-то смысле) объекты, а объекты разных групп должны быть как можно более отличны. Как известно, результаты кластеризации сильно зависят от многих факторов: выбора метрики, способа кластеризации, методов определения количества кластеров. В работе мы рассматриваем новый подход, где кластеризация рассматривается как кооперативная игра с коалиционной структурой и не останавливаемся на вопросе подбора метрики.

Применение кооперативной теории игр для кластеризации данных рассмотрено в [6], но характеристическая функция является супераддитивной и монотонной, как следствие игрокам всегда выгодно объединяться в гранд-коалицию. Этот подход основан на утверждении, что значения вектора Шепли (в игре, рассматриваемой в [6]) будут практически одинаковыми у близких объектов, поэтому можем выбрать в качестве центров кластеров игроков с наибольшим значением вектора Шепли и далее присоединять к ним игроков, у которых значение вектора Шепли не сильно отличается от игроков центра. Процедура итеративная, но эффективная, так как поиск вектора Шепли для характеристической функции, введенной в [6], занимает полиномиальное время. Еще один подход кластеризации, основанный на игре с такой же характеристической функцией, рассмотрен в [5]. В некоторых работах кооперативная теория игр применяется для поиска сообществ в различных коммуникационных сетях [1, 4] и для формирования энергоэффективных кластеров в беспроводной сенсорной сети [3].

2. Модель

Пусть $N = \{1, 2, \ldots, n\}$ – множество объектов, таких, что для любых двух объектов определена функция различия $d(i, j) : N \times N \mapsto [0, 1]$ и функция сходства $f(i, j) : N \times N \mapsto [0, 1]$. Причем для этих функций выполнено свойство симметричности, т.е. $d(i, j) = d(j, i)$ и $f(i, j) = f(j, i)$, $\forall i, j \in N$ и с возрастанием одной функции вторая не

возрастает

$$d(i,j) \geqslant d(k,l) \Longleftrightarrow f(i,j) \leqslant f(k,l) \quad \forall i,j,k,l \in N.$$

В зависимости от практической задачи может быть удобно задать меру различия или меру сходства. Заметим, что достаточно, чтобы была определена хотя бы одна из этих функций, так как можно задать одну функцию через другую, например, $d(i,j) = 1 - f(i,j)$, в таком случае одна из функций имеет определенную интерпретацию в задаче, а вторая является «техническим дополнением» для проведения кластеризации.

Под коалицией S понимается некоторое непустое подмножество множества объектов N, которые объединяются в кластер. Рассмотрим игру (N, v), в которой характеристическая функция игры v определяется следующим образом

$$v(S) = \sum_{\substack{i \in S \\ j \in S \\ i \neq j}} f(i,j) + 2 \sum_{\substack{i \in S \\ j \in N \setminus S}} d(i,j). \tag{2.1}$$

Первое слагаемое этой функции возрастает при присоединении новых объектов к кластеру S, а второе слагаемое может становиться как больше, так и меньше. Такая характеристическая функция будет велика у кластеров, где объекты схожи внутри кластера и в то же время далеки от объектов других кластеров, что соответствует постановке задачи кластеризации.

Пример 2.1. Рассмотрим игру (N, v), где $N = \{1, 2, 3\}$, функции различия и сходства заданы следующими матрицами:

$$d = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 0 & 0.01 & 1 \\ 0.01 & 0 & 0.989 \\ 1 & 0.989 & 0 \end{pmatrix} \end{array}, \quad f = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 1 & 0.99 & 0 \\ 0.99 & 1 & 0.011 \\ 0 & 0.011 & 1 \end{pmatrix} \end{array}.$$

Согласно (2.1) получаем следующие значения характеристической функции v:

$$v(\{1\}) = 2.02, \quad v(\{2\}) = 1.998, \quad v(\{3\}) = 3.978$$
$$v(\{1,2\}) = 5.958, \quad v(\{1,3\}) = 1.998, \quad v(\{2,3\}) = 2.042$$
$$v(\{1,2,3\}) = 2.002.$$

Этот пример демонстрирует нарушение свойств монотонности и супераддитивности функции v.

Утверждение 2.1. *Характеристическая функция $v(\cdot)$, определяемая согласно (2.1), 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 $S \subseteq T$.

Доказательство. Если неравенство (2.1) нарушено в частном случае, значит оно не выполняется и в общем случае, поэтому рассмотрим $S \subset N$, $k \in N$, $k \notin S$, $T = S \cup \{k\}$. Согласно (2.1) имеет место следующее неравенство:

$$v(S \cup \{k\}) - v(S) \geqslant 0. \tag{2.2}$$

Рассмотрение неравенства (2.2) интересно тем, что оно показывает какой вклад (положительный или отрицательный) делает объект в выигрыш коалиции в зависимости от его свойств и свойств коалиции. Рассмотрим левую часть (2.2), по определению характеристической функции (2.1) получаем

$$\sum_{\substack{i\in S\cup\{k\}\\ j\in S\cup\{k\}\\ i\neq j}} f(i,j) + 2\sum_{\substack{i\in S\cup\{k\}\\ j\in N\setminus(S\cup\{k\})}} d(i,j) - \sum_{\substack{i\in S\\ j\in S\\ i\neq j}} f(i,j) - 2\sum_{\substack{i\in S\\ j\in N\setminus S}} d(i,j) =$$

$$= \sum_{\substack{i\in S\\ j\in S\\ i\neq j}} f(i,j) + 2\sum_{i\in S} f(i,k) + 2\sum_{\substack{i\in S\\ j\in N\setminus(S\cup\{k\})}} d(i,j) + 2\sum_{j\in N\setminus(S\cup\{k\})} d(k,j) -$$

$$- \sum_{\substack{i\in S\\ j\in S\\ i\neq j}} f(i,j) - 2\sum_{\substack{i\in S\\ j\in N\setminus(S\cup\{k\})}} d(i,j) - 2\sum_{i\in S} d(i,k) =$$

$$= 2\sum_{i\in S} f(i,k) + 2\sum_{j\in N\setminus(S\cup\{k\})} d(k,j) - 2\sum_{i\in S} d(i,k) =$$

$$= 2\Big(\sum_{i\in S}\big(f(i,k) - d(i,k)\big) + \sum_{i\in N\setminus(S\cup\{k\})} d(i,k)\Big).$$

Таким образом, при предположении монотонности v имеем

$$\forall k \in N, S \subset N \setminus \{k\} \implies \sum_{i \in S} \big(d(i,k) - f(i,k)\big) \leqslant \sum_{i \in N \setminus (S \cup \{k\})} d(i,k). \tag{2.3}$$

Легко можно привести такие $d(i,k)$ и $f(i,k)$, для которых неравенство (2.3) будет нарушаться. В качестве контрпримера можно рассмотреть множества $\{1,2,3\}$ и $\{1,2\}$ в игре (N,v) из примера 2.1. $\square$

Для постановки задачи является важным, что в тривиальном случае, когда сходство между добавленным объектом k и объектами кластера S не меньше, чем расстояние между ними, неравенство (2.3) выполнено.

Утверждение 2.2. *Характеристическая функция $v(\cdot)$, определяемая согласно (2.1), 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 $T \subset N$.

Доказательство. Пусть $S \subset N$, $T \subset N$, $S \cap T = \varnothing$, рассмотрим следующую разность

$$v(S \cup T) - v(S) - v(T),$$

которая должна быть неотрицательной в случае супераддитивности. По определению характеристической функции (2.1) получаем

$$\sum_{\substack{i \in S \cup T \\ j \in S \cup T \\ i \neq j}} f(i,j) + 2 \sum_{\substack{i \in S \cup T \\ j \in N \setminus (S \cup T)}} d(i,j) - \sum_{\substack{i \in S \\ j \in S \\ i \neq j}} f(i,j) - 2 \sum_{\substack{i \in S \\ j \in N \setminus S}} d(i,j) -$$

$$- \sum_{\substack{i \in T \\ j \in T \\ i \neq j}} f(i,j) - 2 \sum_{\substack{i \in T \\ j \in N \setminus T}} d(i,j) = \Big(\sum_{\substack{i \in S \\ j \in S \\ i \neq j}} f(i,j) + \sum_{\substack{i \in T \\ j \in T \\ i \neq j}} f(i,j) + 2 \sum_{\substack{i \in S \\ j \in T}} f(i,j) \Big) +$$

$$+ \Big(2 \sum_{\substack{i \in S \\ j \in N \setminus (S \cup T)}} d(i,j) + 2 \sum_{\substack{i \in T \\ j \in N \setminus (S \cup T)}} d(i,j) \Big) - \sum_{\substack{i \in S \\ j \in S \\ i \neq j}} f(i,j) -$$

$$-\Big(2\sum_{\substack{i\in S\\ j\in N\setminus(S\cup T)}} d(i,j)+2\sum_{\substack{i\in S\\ j\in T}} d(i,j)\Big)-\sum_{\substack{i\in T\\ j\in T\\ i\neq j}} f(i,j)-$$

$$-\Big(2\sum_{\substack{i\in T\\ j\in N\setminus(S\cup T)}} d(i,j)+2\sum_{\substack{i\in T\\ j\in S}} d(i,j)\Big)=2\sum_{\substack{i\in S\\ j\in T}} f(i,j)-4\sum_{\substack{i\in T\\ j\in S}} d(i,j),$$

таким образом функция $v(\cdot)$ супераддитивна тогда и только тогда, когда

$$\forall S,T: S\subset N, T\subset N, S\cap T=\varnothing \implies \tag{2.4}$$

$$\sum_{\substack{i\in S\\ j\in T}} f(i,j)\geqslant 2\sum_{\substack{i\in S\\ j\in T}} d(i,j). \tag{2.5}$$

Напомним, что по определению $f(i,j)\in[0,1]$ и $d(i,j)\in[0,1]$ и с возрастанием одной функции вторая убывает. Очевидно, что в общем случае условие (2.4)-(2.5) не выполнено. В качестве контрпримера можно рассмотреть множества $\{1\},\{3\}$ и $\{1,3\}$ в игре (N,v) из примера 2.1. □

Из (2.4)-(2.5) следует, что объединять 2 кластера выгодно только при условии, что сходство объектов хотя бы в 2 раза больше, чем расстояние между ними.

Отсутствие свойства супераддитивности говорит о том, что в общем случае игрокам не выгодно объединяться в гранд-коалицию, что естественно для задачи кластеризации.

3. Игры с коалиционной структурой

В связи с отсутствием свойства супераддитивности у характеристической функции (2.1) имеет смысл рассматривать игры с коалиционной структурой, в нашей задаче коалиционную структуру можно интерпретировать как возможный вариант кластеризации. Таким образом, задача кластеризации сводится к поиску определенной «наиболее выигрышной» коалиционной структуры. Далее будем рассматривать игру (N,v,π) со множеством игроков N, характеристической функцией v и коалиционной структурой π.

Определение 3.1. *Коалиционная структура π – это разбиение множества всех игроков N следующим образом: $\pi = \{B_1, \ldots, B_m\}$, причем*

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

Определение 3.2. *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 3.3. *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.

Для того, чтобы сформулировать определение устойчивой коалиционной структуры [6], введем следующие обозначения: разбиение π_{-B_k}, такое что $\pi_{-B_k} = \pi \setminus B_k$; разбиение π', которое 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 $B \in \pi_{-B(k)} \cup \varnothing$.

Определение 3.4. *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':*

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

Такое определение устойчивой коалиционной структуры соответствует концепции равновесия Нэша [6]. 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 [7].

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.*

4. Существование устойчивых коалиционных структур

Утверждение 4.1. *В игре (N, v, π) с коалиционной структурой π и характеристической функцией v вида (2.1) компоненты вектора*

Ауманна–Дрезе ϕ^π заданы 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 $|S| = s$.

Доказательство. В доказательстве утверждения 2.1 показано, что

$$v(S\cup\{k\})-v(S) = 2\left(\sum_{i\in S}\big(f(i,k)-d(i,k)\big)+\sum_{i\in N\setminus(S\cup\{k\})} d(i,k)\right), \tag{4.1}$$

тогда, переобозначив $S\cup\{k\} := S$ и $S := S\setminus\{k\}$, по определению вектора Ауманна–Дрезе получаем

$$\phi_k^\pi = \sum_{\substack{S\subseteq B(k)\\ S\ni k}} \frac{(b-s)!(s-1)!}{b!}\Big[v(S)-v(S\setminus\{k\})\Big] \overset{(4.1)}{=}$$

$$\overset{(4.1)}{=} \sum_{\substack{S\subseteq B(k)\\ S\ni k}} \frac{2(b-s)!(s-1)!}{b!}\Big[\sum_{i\in S\setminus\{k\}}\big(f(i,k)-d(i,k)\big)+\sum_{i\in N\setminus S} d(i,k)\Big].$$

□

Из утверждения 4.1 можно получить условия неотрицательности компонент вектора Ауманна–Дрезе.

Следствие 4.1. *$\phi_k^\pi \geqslant 0$ тогда и только тогда, когда выполнено следующее неравенство*

$$\sum_{\substack{S\subseteq B(k)\\ S\ni k}} \frac{2(b-s)!(s-1)!}{b!}\Big[\sum_{i\in S\setminus\{k\}}\big(f(i,k)-d(i,k)\big)+\sum_{i\in N\setminus S} d(i,k)\Big] \geqslant 0. \tag{4.2}$$

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 4.2. *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* $B \cap B(k) = \varnothing$, $B \in \pi \cup \varnothing$, $B(k) \in \pi$.

Доказательство. Следует из определения 3.4 и утверждения 4.1. □

Замечание 4.1. В игре (N, v, π) не всегда существует устойчивая относительно вектора Ауманна–Дрезе коалиционная структура.

Утверждение 4.3. *В игре (N, v, π) с коалиционной структурой π и характеристической функцией v вида (2.1) компоненты ES-вектора ξ^π заданы формулой:*

$$\xi_k^\pi = 2 \sum_{j \in N \setminus \{k\}} d(k,j) + \frac{1}{b} \Big(\sum_{\substack{i \in B(k) \\ j \in B(k) \\ i \neq j}} \big(f(i,j) - 2d(i,j)\big) \Big), \tag{4.3}$$

где $|B(k)| = b$.

Доказательство. По определению ES-вектора и характеристической функции v:

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

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

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

$$-2\sum_{\substack{i\in B(k)\\ j\in N\setminus B(k)}} d(i,j)-2\sum_{\substack{i\in B(k)\\ j\in B(k)\\ i\neq j}} d(i,j)\Big)=$$

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

□

Из утверждения 4.3 можно получить условия неотрицательности компонент ES-вектора.

Следствие 4.2. $\xi_k^\pi \geqslant 0$ *тогда и только тогда, когда выполнено следующее неравенство*

$$2\sum_{j\in N\setminus\{k\}} d(k,j)\geqslant \frac{1}{b}\Big(\sum_{\substack{i\in B(k)\\ j\in B(k)\\ i\neq j}} \big(2d(i,j)-f(i,j)\big)\Big). \tag{4.4}$$

Можно заметить, что в случае объединения в кластер сильно схожих объектов выполняется условие неотрицательности компонент вектора (4.4). Под сильно схожими объектами $i,j\in N$ здесь понимаются такие, что выполняется $f(i,j)\geqslant 2d(i,j)$.

Утверждение 4.4. *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* $B\cap B'=\varnothing$, $B'\in\pi\cup\varnothing$, $B\in\pi$.

Доказательство. Следует из определения 3.4 и утверждения 4.3. □

Следствие 4.3. *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 4.2. В игре (N, v, π) не всегда существует устойчивая относительно вектора ES-вектора структура.

Пример 4.1. Пусть заданы точки на плоскости, координаты которых приведены в таблице 1. Эти точки являются объектами кластеризации и мы хотим разделить их на группы по геометрическому расстоянию, на рисунке 4.1 показан желаемый результат.

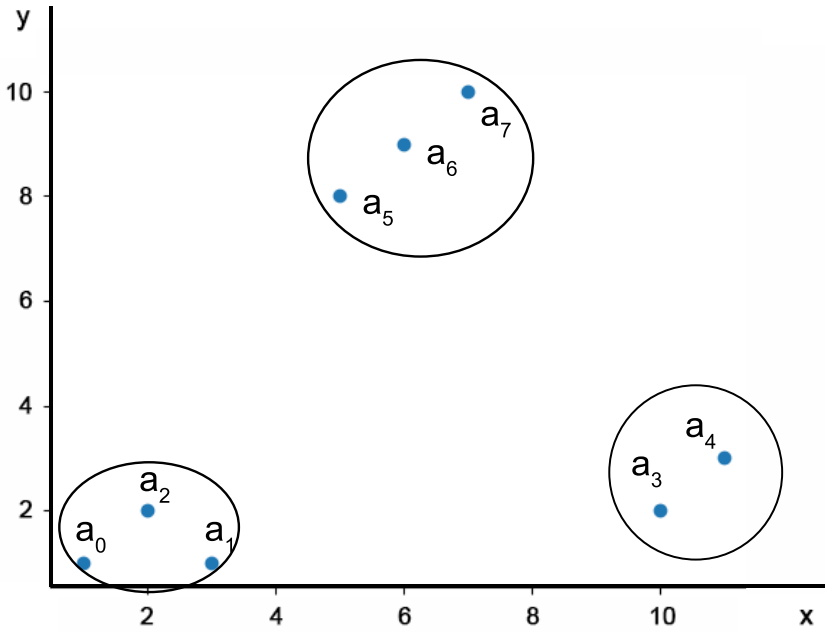

Рисунок 1. Кластеризация точек на плоскости

Таблица 1.

№	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7
x	1	3	2	10	11	5	6	7
y	1	1	2	2	3	8	9	10

Будем считать, что $d(a_i, a_j)$, где $i, j \in N = \{0, 1, \ldots, 7\}$ вычисляется как евклидово расстояние, нормированное на расстояние между самыми далекими точками

$$d(a_i, a_j) = \frac{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{d_{max}},$$

$$d_{max} = \max_{i,j \in N} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2},$$

где x_i, y_i – это координаты точек, $a_i = (x_i, y_i)$. Сходство между точками определим как $f(a_i, a_j) = 1 - d(a_i, a_j)$. В матрице ниже над диагональю расположены значения расстояния между точками, а под диагональю – сходства.

$$A = \begin{array}{c} \\ a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \end{array} \begin{array}{c} \begin{array}{cccccccc} a_0 & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 \end{array} \\ \begin{pmatrix} - & 0.18 & 0.13 & 0.84 & 0.94 & 0.75 & 0.87 & 1 \\ 0.82 & - & 0.13 & 0.65 & 0.76 & 0.67 & 0.79 & 0.91 \\ 0.87 & 0.87 & - & 0.74 & 0.84 & 0.62 & 0.75 & 0.87 \\ 0.16 & 0.35 & 0.26 & - & 0.13 & 0.72 & 0.75 & 0.79 \\ 0.06 & 0.24 & 0.16 & 0.87 & - & 0.72 & 0.72 & 0.75 \\ 0.25 & 0.33 & 0.38 & 0.28 & 0.28 & - & 0.13 & 0.26 \\ 0.13 & 0.21 & 0.25 & 0.25 & 0.28 & 0.87 & - & 0.13 \\ 0.00 & 0.09 & 0.13 & 0.21 & 0.25 & 0.74 & 0.87 & - \end{pmatrix} \end{array}.$$

Таким образом получаем игру (N, v), где $N = \{0, 1, \ldots, 7\}$ и v задана в (2.1). Приведем значения характеристической функции для множеств из одного игрока (с точностью до двух знаков после запятой):

$$v(\{0\}) = 9.43, \quad v(\{1\}) = 8.21, \quad v(\{2\}) = 8.15, \quad v(\{3\}) = 9.24,$$
$$v(\{4\}) = 9.73, \quad v(\{5\}) = 7.74, \quad v(\{6\}) = 8.27, \quad v(\{7\}) = 9.42.$$

Численный эксперимент, в котором рассмотрены все возможные разбиения множества N показал, что коалиционная структура

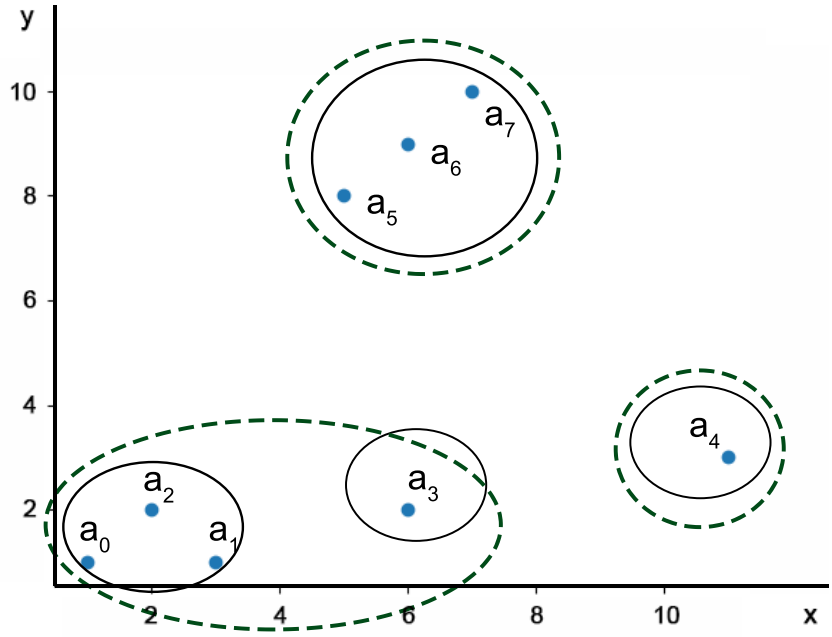

Рисунок 2. Кластеризация точек на плоскости: сплошная линия – первый вариант разбиения, пунктир – второй вариант разбиения

$\pi^* = \{\{0,1,2\},\{3,4\},\{5,6,7\}\}$ является устойчивой относительно вектора Ауманна–Дрезе. Дележ относительно вектора Ауманна–Дрезе (с точностью до двух знаков после запятой) равен

$$\phi^{\pi^*} = (10.48,\ 9.26,\ 9.37,\ 9.84,\ 10.33,\ 8.57,\ 9.48,\ 10.24),$$

можно увидеть, что необходимое условие (теорема 3.1) удовлетворено, выполнение достаточного условия проверено с помощью реализации программы. Можно заметить, что компоненты вектора Ауманна–Дрезе больше у тех игроков, которые имеют наибольшее суммароное расстояние до других игроков. И наоборот, меньше у тех игроков, которые имеют наименьшее суммарное расстояние до других игроков.

Эта же коалиционная структура является устойчивой относительно ES-вектора, дележ (с точностью до двух знаков после запятой) равен

$$\xi^{\pi^*} = (10.53,\ 9.31,\ 9.25,\ 9.84,\ 10.33,\ 8.7,\ 9.22,\ 10.37).$$

Компоненты двух векторов очень близки по значениям. Однако в некоторых случаях вектора могут значительно различаться и даже давать различные решения. Рассмотрим менее однозначный пример (рисунок 4.1), который получается изменением координат точки с $a_3 = (10, 2)$ на $a_3 = (6, 2)$. В таком случае коалиционная структура $\hat{\pi} = \{\{0, 1, 2\}, \{3\},\ \{4\}, \{5, 6, 7\}\}$ является устойчивой относительно вектора Ауманна–Дрезе, соответствующий дележ равен

$$\phi^{\hat{\pi}} = (9.74,\ 8.54,\ 8.62,\ 7.11,\ 10.41,\ 8.25,\ 9.29,\ 10.15).$$

Относительно ES-вектора устойчива коалиционная структура $\tilde{\pi} = \{\{0, 1, 2, 3\}, \{4\}, \{5, 6, 7\}\}$, соответствующий дележ равен

$$\xi^{\tilde{\pi}} = (9.32,\ 8.11,\ 8.04,\ \mathbf{7.74},\ 10.41,\ 8.38,\ 9.03,\ 10.28).$$

Примечательно то, что суммарный выигрыш $\sum\limits_{k=1}^{7} \xi_k^{\tilde{\pi}}$ коалиции $\tilde{\pi}$ равен 71.34, в то время как суммарный выигрыш $\sum\limits_{k=1}^{7} \xi_k^{\hat{\pi}}$ коалиции $\hat{\pi}$ составляет 72.14. Однако такая коалиция не является устойчивой относительно ES-вектора, так как игроку 3 выгоднее присоединиться к коалиции $\{0, 1, 2\}$:

$$\xi^{\hat{\pi}} = (9.74,\ 8.54,\ 8.62,\ \mathbf{7.11},\ 10.4,\ 8.25,\ 9.29,\ 10.15).$$

Действительно, после присоединения игрока 3 к коалиции $\{0, 1, 2\}$ среднее сходство объектов коалиции становится меньше, что приводит к уменьшению выигрыша. Из второго слагаемого 4.3 можно увидеть, что большие коалиции являются более предпочтительными относительно ES-вектора, что объясняет результат, полученный выше.

5. Заключение

В работе представлена новая теоретико-игровая постановка задачи кластеризации. Рассматривается игра с характеристической функцией, которая принимает большие значения при формировании «правильного» (с точки зрения определенной метрики) кластера. Такая

функция не обладает свойством супераддитивности, поэтому рассматривается игра с коалиционной структурой. Наличие стабильной, относительно выбранного принципа оптимальности, коалиционной структуры говорит о хорошей разделимости кластеров, т.е. о существовании четкой кластеризации. В дальнейшем возможно рассмотрение построения нечеткой кластеризации при отсутствии стабильной коалиционной структуры. К достоинствам метода можно отнести отсутствие необходимости предопределять количество кластеров, к недостаткам – экспоненциальный рост сложности, поэтому в будущем планируется разработать более эффективный алгоритм для поиска стабильных коалиционных структур.

СПИСОК ЛИТЕРАТУРЫ

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

2. 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.

3. Afsar M. *Energy-Efficient Coalition Formation in Sensor Networks: a Game-Theoretic Approach.* Computing Research Repository. 2015. abs/1512.08019.

4. Avrachenkov K.E., Kondratev A.Yu., Mazalov V.V. *Cooperative Game Theory Approaches for Network Partitioning* // International Computing and Combinatorics Conference. 2017. LNCS 10392. P. 591–602.

5. Garg V.K., Narahari Y. *Novel biobjective clustering (BiGC) based on cooperative game theory* // IEEE Transactions on Knowledge and Data Engineering. 2013. Vol. 25. Is. 5. P. 1070–1082.

6. Dhamal S., Bhat S., Anoop K.R., & Embar V.R. *Pattern Clustering using Cooperative Game Theory* // Computing Research Repository.

2012. abs/1201.0461. P. 653–658.

7. 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.

APPLYING COOPERATIVE GAME THEORY WITH COALITIONAL STRUCTURE FOR DATA CLUSTERING

Vladimir M. Bure, Saint Petersburg State University, Dr.Sc., professor (v.bure@spbu.ru),
Kseniya Yu. Staroverova, Saint Petersburg State University, postgraduate student (st016234@student.spbu.ru).

Abstract: In the work a cooperative game where distance or similarity of players may be defined is considered. A characteristic function is defined in such a way that it is high for such coalitions which consist of more similar objects than objects from other coalitions. We consider the function which may not be superadditive that is why not only the grand coalition but smaller ones can be formed and considering a game with coalitional structure is reasonable. Therefore, we have a natural transformation from a cooperative game to a clustering problem. Several single-valued cooperative solution concepts are considered with this type of a characteristic function and stability conditions are found.

Keywords: coalition, stable coalitional structure, clustering, Shapley value, Aumann-Dreze value, ES-value.