

Обобщенный алгоритм формирования классов структурных состояний информационных систем¹

А.Н. Павлов, Б.В. Соколов, С.А. Осипенко

Аннотация. Предлагается обобщенный алгоритм многокритериальной кластеризации многоструктурных состояний катастрофоустойчивых информационных систем (КАИС). Алгоритм позволяет выделить эквивалентные структурные состояния КАИС и ядра кластеров, которые используются для идентификации структурных состояний реконфигурируемой системы.

Ключевые слова: многокритериальная кластеризация, катастрофоустойчивая информационная система, многоструктурное состояние, реконфигурация.

В современных условиях спектр угроз экономической, физической и информационной безопасности, а также перечень уязвимостей технической инфраструктуры бизнеса (производственных процессов) и, в частности, информационных систем (ИС), постоянно растет. В этих условиях обеспечение непрерывности бизнес-процессов и повышение катастрофоустойчивости соответствующих бизнес-систем являются одним из важнейших стратегических направлений развития любой компании. В этих условиях исследовать и решать задачи обеспечения катастрофоустойчивости ИС (КАИС) необходимо уже в рамках междисциплинарного подхода, интерпретируя их как задачи управления структурной динамикой. Широкое распространение на практике получил такой вариант управления структурами КАИС, как реконфигурация.

Реконфигурация КАИС – это процесс изменения структуры КАИС в целях сохранения, восстановления (повышения) уровня работо-

способности КАИС либо обеспечения минимального снижения уровня эффективности применения КАИС при деградации её функций.

Одной из важнейших прикладных задач, возникающих в этом контексте, является задача предварительной кластеризации возможных многоструктурных состояний КАИС и выделение типовых структур кластеров. Решение предлагаемых задач позволит повысить оперативность определения структурного состояния реконфигурируемой КАИС и предварительного анализа ее свойств.

Формально постановка задачи кластеризации множества многоструктурных состояний КАИС сводится к следующему: известно множество многоструктурных состояний $S = \{S_1, S_2, \dots, S_n\}$, каждое многоструктурное состояние характеризуется конечной совокупностью показателей (признаков) $\vec{F} = (F_1, F_2, \dots, F_m)$, с помощью которых оцениваются различные аспекты функционирования КАИС (например: безопасность, надежность, устойчивости, гибкость и т.п.), при

¹ Исследования, выполненные по данной тематике, проводились при финансовой поддержке Российского фонда фундаментальных исследований (гранты 07-07-00169, 06-07-89242, 08-08-00403, 09-07-00066а), Отделения нанотехнологий и информационных технологий РАН (проект №О-2.5/03).

этом рассматриваемые показатели оценивают различные (техническую, топологическую, технологическую и др.) структуры КАИС, т.е. характеризуют многоструктурное состояние КАИС. Для классификации многоструктурных состояний S введем различные меры сходства и включения, порождающие отношения на исследуемом множестве. Под мерой сходства будем понимать неотрицательную вещественную функцию $C(S_l, S_k)$, обладающую следующими свойствами:

$$0 \leq C(S_l, S_k) \leq 1, C(S_l, S_k) = 1, \\ \text{если } l = k, C(S_l, S_k) = C(S_k, S_l). \quad (1)$$

Свойствами (1) обладает, в частности, континуум эквивалентных мер [2,3], представляемых формулой

$$C(S_l, S_k) = \frac{2\mu(S_l \cap S_k)}{(1 + \lambda)[\mu(S_l) + \mu(S_k)] - 2\lambda\mu(S_l \cap S_k)} \quad (2)$$

где $-1 \leq \lambda < \infty$; $\mu(S_i)$ – мощность множества признаков, удовлетворяющих S_i структуре

В [2, 3] приведены различные меры сходства структур, предназначенные для обработки качественных и количественных признаков.

При этом
$$\mu(S_l) = \sum_{j=1}^m F_{lj}, \mu(S_l \cap S_k) = \\ = \sum_{j=1}^m \min\{F_{lj}, F_{kj}\}, \mu(S_l \cup S_k) = \sum_{j=1}^m \max\{F_{lj}, F_{kj}\}.$$

Если сравнение структур осуществляется по качественным признакам, то $F_{ij} \in \{0, 1\}$ для количественных признаков F_{ij} принимают значения либо в установленной ранговой шкале, либо в интервале $[0, 1]$.

В отличие от мер сходства мера включения отражает различную степень включения одной структуры в другую по сравниваемым признакам и определяет, какая структура более оригинальна, а какая – более типична среди множества S . Меры включения структуры S_l в структуру S_k и наоборот определяются следующим образом:

$$V(S_l, S_k) = \frac{\mu(S_l \cap S_k)}{\mu(S_k)}, V(S_k, S_l) = \frac{\mu(S_l \cap S_k)}{\mu(S_l)}. \quad (3)$$

Если $V(S_l, S_k) > V(S_k, S_l)$, то S_l структура более оригинальна, а S_k более типовая.

С использованием мер сходства можно построить отношение иерархии, которое позволит выделить кластеры эквивалентных структур из исследуемого множества S , после чего для каждого кластера структур на основе мер включения можно выделить наиболее типовой элемент.

Отношение иерархии или иерархическое разбиение структур определяется следующим образом [4].

Здесь под разбиением множества S на r кластеров понимается множество подмножеств

$$R = \{H_q \mid H_q \subseteq S : \bigcup_{q=1}^r H_q = S; H_q \cap H_l = \emptyset \forall q \neq l\}.$$

Тогда иерархическое разбиение (отношение иерархии) множества S есть множество разбиений $T = \{R_\alpha, \alpha = 0, 1, 2, \dots\}$, удовлетворяющее

следующему условию вложенности разбиений: $\forall R_{\alpha_1}, R_{\alpha_2}$, если $\alpha_1 < \alpha_2$, то $\forall H_q \in R_{\alpha_1} \exists H_p \in R_{\alpha_2}$:

$H_q \subseteq H_p$ и $\forall H_r \in R_{\alpha_2} \setminus \{H_p\} H_q \cap H_r = \emptyset$. Так, например, в иерархическое разбиение входят разбиения, характеризующие предельные случаи,

$$R_0 = \{H_q \mid H_q = \{S_q\}, q = 1, 2, \dots, n\} \quad \text{и}$$

$$R_n = \{H_1 \mid H_1 = \{S_1, S_2, \dots, S_n\}\}.$$

При этом если задано разбиение, то структуры, входящие в один и тот же кластер, являются неразличимыми (эквивалентными).

При практической реализации иерархических классификаций строятся дендрограммы, являющиеся графическим способом изображения системы, что делает наглядной структуру иерархического отношения.

Для построения иерархического разбиения множества структур S могут быть использованы методы сгущения кластеров [3], пересчет значений сходства которых можно осуществлять по следующим формулам:

$$C(H_j, H_l) = \alpha_q C(H_j, H_q) + \alpha_l C(H_j, H_l) + \\ + \beta C(H_q, H_l) + \nu [C(H_j, H_q) - C(H_j, H_l)] \quad (4)$$

Параметры $\alpha_q, \alpha_l, \beta, \nu$ задают вид метода пересчета меры сходства кластеров: метод минимума, максимума, медианы, центроидный метод, метод Уорда.

В первой части обобщенного алгоритма осуществляется построение дендрограммы, используя которую можно формировать различные разбиения исходного множества S с требуемой степенью эквивалентности (γ) входящих в кластеры структур, т.е., варьируя параметр γ , можно получать кластеры различной степени эквивалентности.

Во второй части обобщенного алгоритма для выявления наиболее типовых структур кластера используем меру включения. На основе рассчитанных значений мер включения для структур одного кластера строится матрица размерностью $p \times p$, где p - мощность рассматриваемого кластера. Для данной матрицы находим собственный вектор, соответствующий максимальному собственному числу матрицы. Максимальное значение среди компонент найденного вектора соответствует наиболее типовому элементу кластера, а минимальное значение – наиболее оригинальной структуре [3].

Продемонстрируем работу обобщенного алгоритма формирования классов структурных состояний для случая, в котором рассмотрим 11 простых сетевых структур $S = \{S_1, S_2, \dots, S_{11}\}$, представленных на Рис. 1.

Функционирование рассматриваемых структур будем оценивать по следующим трем признакам: F_1 - структурная надежность (устойчивость), F_2 - сложность сети, F_3 - степень децентрализации структуры (структурная гибкость).

Остановимся подробнее на указанных признаках.

Для оценивания структурной надежности (устойчивости) сети в зависимости от ее построения применим подход, предложенный в работах [5, 6]. Как правило, задача структурного анализа сводится к построению соответствующих структурных функций - функций надежности КАИС, функций безопасности, функций живучести, функций работоспособности, минимальных сечений отказов, путем ортогонализации функций алгебры логики (ФАЛ), замещения логических аргументов в ФАЛ вероятностями их истинности и логических операций арифметическими. Теоретической основой указанного подхода является математический аппарат логико-вероятностных исчислений [5], применение которого позволяет

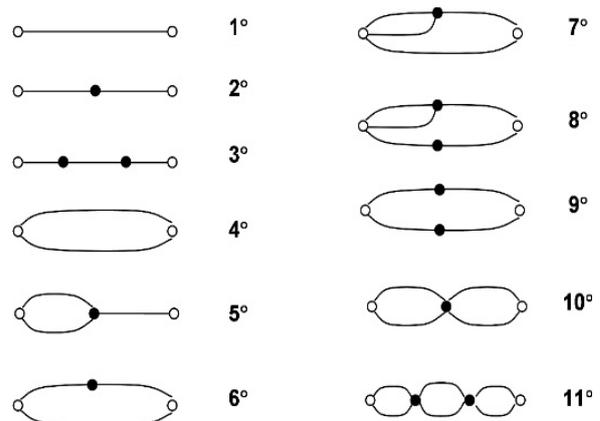


Рис. 1 Сетевые структуры

количественно оценить надежность структуры в целом в зависимости от вероятности надежной работы входящих в нее элементов.

Вероятность надежной работы сети определяется составом ее элементов и структурой их взаимосвязей. Расчет вероятности надежной работы сети можно осуществлять по следующей формуле [5]:

$$R(P_1, P_2, \dots, P_n) = P\{\bigvee_{i=1}^d \Pi_i\} = \sum_i P(\Pi_i) - \sum_i \sum_j P(\Pi_i \wedge \Pi_j) + \sum_i \sum_j \sum_k P(\Pi_i \wedge \Pi_j \wedge \Pi_k) - \dots + (-1)^{d-1} P(\Pi_1 \wedge \Pi_2 \wedge \dots \wedge \Pi_d); \quad (5)$$

где $\{\Pi_l, l = 1, \dots, d\}$ - множество кратчайших путей успешного функционирования сети. Приведенную формулу можно переписать в следующем виде

$$R(P_1, P_2, \dots, P_n) = \sum_{l=1}^v (-1)^{l-1} \sum_{i=1}^{C_v^l} \prod_{q \in W_i(\Pi, l)} P_q; \quad (6)$$

где v - количество кратчайших путей, C_v^l - число сочетаний из v по l , $W_i(\Pi, l)$ - подмножество номеров элементов сети из $\{1, 2, \dots, n\}$, получающихся при i -ом варианте объединения элементов различных l путей из множества всех путей Π , P_q - вероятность надежной работы q -го элемента.

Ценность полученного полинома $R(P_1, P_2, \dots, P_n)$ от n переменных P_i состоит в том, что он позволяет количественно оценить вероятность надежной работы сети при разных P_i . При условии равновероятной надежной

работы всех элементов $P_1 = P_2 = \dots = P_n = P$ из выражений (6) получим полином $R(P)$ от одной переменной. Полином $R(P)$, как показано в [6], отображает некоторые интересные и очень полезные свойства сети, связанные с ее надежностью.

В частности, полином $R(P)$ монотонно возрастает и всегда находится в пределах $0 \leq R(P) \leq 1$, при этом $R(0) = 0$, $R(1) = 1$. Чем ближе график полинома к прямой $R(P) = 1$, тем надежнее сеть. Поэтому в качестве интегрального показателя структурной надежности сети при одинаковой надежности элементов предлагается в [5] использовать

$$F_1 = \int_0^1 R(P) dP. \quad (7)$$

Произведем построение полиномов надежности приведенных на Рис.1 структур и рассчитаем показатель $F_1(S_i)$, $i = 1, 2, \dots, 11$.

Полиномы, характеризующие надежность этих структур, представлены следующими выражениями

$$\begin{aligned} R_1(P) &= P, R_2(P) = P^2, R_3(P) = P^3, \\ R_4(P) &= 2P - P^2, R_5(P) = 2P^2 - P^3, \\ R_6(P) &= P + P^2 - P^3, \\ R_7(P) &= P + 2P^2 - 3P^3 + P^4, \\ R_8(P) &= 3P^2 - P^3 - 2P^4 + P^5, \\ R_9(P) &= 2P^2 - P^4, R_{10}(P) = 4P^2 - 4P^3 + P^4, \\ R_{11}(P) &= 8P^3 - 12P^4 + 6P^5 - P^6. \end{aligned}$$

Соответственно, показатель структурной надежности $F_1(S_i) = \int_0^1 R_i(P) dP$, $i = 1, 2, \dots, 11$,

будет иметь следующие значения

$$\begin{aligned} F_1(S_1) &= 0.5, F_1(S_2) = 0.33, F_1(S_3) = 0.25, \\ F_1(S_4) &= 0.67, F_1(S_5) = 0.42, F_1(S_6) = 0.58, \\ F_1(S_7) &= 0.62, F_1(S_8) = 0.52, F_1(S_9) = 0.47, \\ F_1(S_{10}) &= 0.53, F_1(S_{11}) = 0.46. \end{aligned}$$

Оценка сложности сети определяется как отношение суммарного числа дуг сети ($L(S_j)$)

к максимально возможному числу дуг в n -вершинном графе $F_2(S_j) = \frac{L(S_j)}{n(n-1)}$.

Для нашего случая получаем следующие результаты:

$$\begin{aligned} F_2(S_1) &= 0.5, F_2(S_2) = 0.333, F_2(S_3) = 0.25, \\ F_2(S_4) &= 0.333, F_2(S_5) = 0.25, F_2(S_6) = 0.5, \\ F_2(S_7) &= 0.3, F_2(S_8) = 0.233, F_2(S_9) = 0.333, \\ F_2(S_{10}) &= 0.19, F_2(S_{11}) = 0.133. \end{aligned}$$

Степень централизации системы довольно сильно определяет основные функциональные свойства системы, а также ее эволюционные перспективы в конкретной среде. В децентрализованной системе нет ярко выраженного главного звена (подсистемы) и главных связей. Другими словами, отсутствует "непререкаемый авторитет" или ключевой элемент. Поэтому решение о том, как изменить свои связи, каждая подсистема принимает самостоятельно. Отсюда, система в целом относительно легко может изменить свою структуру в зависимости от ситуации и своих собственных критериев правильности поведения, т.е. обладает высокими адаптационными способностями (структурной гибкостью).

Для оценивания структурной гибкости сети будем использовать показатель ее децентрализации. Один из них, связанный с введением так называемого индекса центральности, сводится к подсчету по матрице смежности $M_c = \|a_{kj}^c\|_{n \times n}$ суммарных чисел входящих и выходящих дуг для каждой вершины $v(x_j) = \sum_k a_{kj}^c + \sum_k a_{jk}^c$ ($j = 1, 2, \dots, n$), нахождению наибольшей из них $v(x_s)$, определению разностей между $v(x_s)$, $v(x_j)$ ($j = 1, 2, \dots, n$) и последующему суммированию этих разностей. Нетрудно подсчитать, что максимально возможное значение этой суммы будет $2(n-1)(n-2)$. Тогда индекс децентрализованности определится следующим соотношением:

$$F_3 = \max \left\{ 0, 1 - \frac{1}{2(n-1)(n-2)} \sum_{j=1}^n [v(x_s) - v(x_j)] \right\} \quad (8)$$

Проведенные вычисления для рассматриваемых структур дали следующий результат $F_3(S_1) = 1, F_3(S_2) = 0, F_3(S_3) = 0.677,$
 $F_3(S_4) = 1, F_3(S_5) = 0.583, F_3(S_6) = 1,$
 $F_3(S_7) = 0.75, F_3(S_8) = 0.8, F_3(S_9) = 1,$
 $F_3(S_{10}) = 0.6, F_3(S_{11}) = 0.778.$

Перейдем к расчету мер сходства и включения для множества сетевых структур $S = \{S_1, S_2, \dots, S_{11}\}$ по указанным признакам F_1, F_2, F_3 (Табл. 1)

Значение меры сходства вычисляем по формуле (2) с параметром $\lambda = 3$ (формула Сокала-Снита)

$$C(S_l, S_k) = \frac{\mu(S_l \cap S_k)}{2\mu(S_l) + 2\mu(S_k) - 3\mu(S_l \cap S_k)}$$

Используя полученную матрицу сходства структур, осуществляем построение иерархического разбиения структур на кластеры. При этом воспользуемся формулой пересчета значений сходства (4) с параметрами по методу максимума $\alpha_q = \alpha_t = 0.5, \beta = 0, \nu = 0.5,$

$$C(H_j, H_l) = \max\{C(H_j, H_q), C(H_j, H_l)\}.$$

Дендрограмма иерархического разбиения представлена на Рис. 2.

Для степени эквивалентности $0,72 < \gamma \leq 0,75$ получено разбиение структур на 5 кластеров

$$R_\gamma = \{H_1 = \{S_1, S_4, S_6, S_9\}, H_2 = \{S_7, S_8, S_{11}\},$$

$$H_3 = \{S_5, S_{10}\}, H_4 = \{S_3\}, H_5 = \{S_2\}\}.$$

Для каждого из построенных кластеров определим наиболее типовую структуру среди входящих в него элементов. С этой целью для каждого из кластеров построим матрицу включения структур этого кластера с использованием формулы (3).

$$V(S_l, S_k) = \frac{\mu(S_l \cap S_k)}{\mu(S_k)} = \frac{\sum_{j=1}^m \min\{F_{lj}, F_{kj}\}}{\sum_{j=1}^m F_{kj}},$$

$$V(S_k, S_l) = \frac{\mu(S_l \cap S_k)}{\mu(S_l)} = \frac{\sum_{j=1}^m \min\{F_{lj}, F_{kj}\}}{\sum_{j=1}^m F_{lj}}.$$

Табл. 1

		Показатели		
		Надежность	Сложность	Гибкость
структуры	S ₁	0,5	0,5	1
	S ₂	0,33	0,333	0
	S ₃	0,25	0,25	0,677
	S ₄	0,66	0,333	1
	S ₅	0,42	0,25	0,583
	S ₆	0,58	0,5	1
	S ₇	0,62	0,3	0,75
	S ₈	0,52	0,233	0,8
	S ₉	0,47	0,333	1
	S ₁₀	0,53	0,19	0,6
	S ₁₁	0,46	0,133	0,778

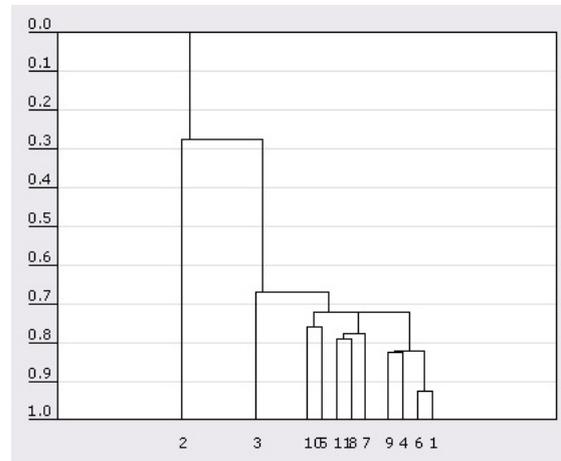


Рис. 2. Дендрограмма

Матрицы включения, геометрические средние и собственные вектора, соответствующие максимальному собственному числу, кластеров разбиения

$$R_\gamma = \{H_1 = \{S_1, S_4, S_6, S_9\}, H_2 = \{S_7, S_8, S_{11}\},$$

$$H_3 = \{S_5, S_{10}\}, H_4 = \{S_3\}, H_5 = \{S_2\}\}.$$

приведены в Табл. 2.

Таким образом, для кластера H_1 из структур $\{S_1, S_4, S_6, S_9\}$ (Рис. 3) наиболее типовой является структура S_4 .

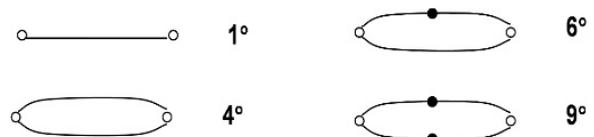


Рис. 3

Табл. 2

Матрица включения кластера $H_1 = \{S_1, S_4, S_6, S_9\}$				Геометрическое среднее	Собственный вектор
1	0,752	0,775	0,798	0,825718	0,218129
1,000	1	1,000	1,000	1	0,264169
1,000	0,970	1	1,000	0,992395	0,26216
0,985	0,929	0,957	1	0,967347	0,255543
Матрица включения кластера $H_2 = \{S_7, S_8, S_{11}\}$				Геометрическое среднее	Собственный вектор
1	0,922	1,000		0,973451	0,342607
0,953	1	1,000		0,984039	0,346333
0,844	0,817	1		0,883817	0,31106
Матрица включения кластера $H_3 = \{S_5, S_{10}\}$				Геометрическое среднее	Собственный вектор
1		0,601		0,77544	0,436759
1,000		1		1	0,563241

Для кластера H_2 из структур $\{S_7, S_8, S_{11}\}$ (Рис.4) наиболее типовая - S_8 .

Для кластера H_3 из структур $\{S_5, S_{10}\}$ (Рис.5) наиболее типовая - S_{10} .

Заключение

В результате проведенных исследований удалось построить алгоритм формирования классов эквивалентных структурных состояний КАИС, который позволяет на этапе применения ИС существенно сократить время на поиск требуемого структурного состояния при их реконфигурации в условиях возникающих нештатных и аварийных событий.

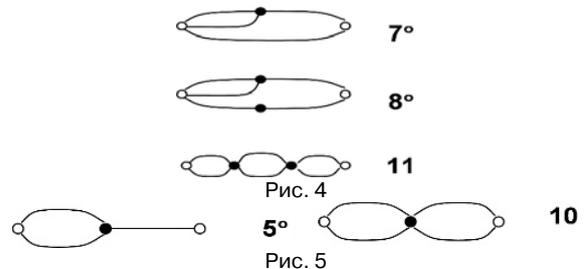
Литература

1. Охтилев М.Ю., Соколов Б.В., Юсупов Р.М. Интеллектуальные технологии мониторинга и управления структурной динамикой сложных технических систем. – М.: Наука, 2006. – 410 с.

Соколов Борис Владимирович. Заместитель директора по научной работе Санкт-Петербургского института информатики и автоматизации РАН (СПИИ РАН). Окончил Военно-инженерную Краснознаменную академию имени А.Ф.Можайского в 1974 году. Доктор технических наук, профессор, Заслуженный деятель науки РФ. Автор более 200 печатных трудов и 5 монографий. Область научных интересов: моделирование сложных систем и принятие решений. E-mail: sokol@iias.spb.su.

Павлов Александр Николаевич. Старший научный сотрудник Санкт-Петербургского института информатики и автоматизации РАН (СПИИ РАН). Кандидат технических наук, доцент. Окончил Ленинградский государственный университет имени А.А.Жданова в 1979 году. Автор более 70 печатных трудов. Область научных интересов: системный анализ, принятие решений в условиях существенной неопределенности и многокритериальности. E-mail: pavlov62@list.ru.

Осипенко Сергей Александрович. Адьютант Военно-космической академии имени А.Ф.Можайского. Окончил в Военно-космическую академию имени А.Ф.Можайского 2003 году. Автор 5 печатных трудов. Область научных интересов: принятие решений в условиях неопределенности. E-mail: fon_vakano@mail.ru.



2. Андреев В.Л. Классификационные построения в экологии и систематике. - М.: Наука, 1980. - 142 с.
3. Андрейчиков А.В., Андрейчикова О.Н. Анализ, синтез, планирование решений в экономике: Учебник. – М.: Финансы и статистика, 2004. – 464 с.
4. Павлов А.Н., Соколов Б.В. Системный анализ организационных структур управления: учеб. пособие / ГУАП – СПб., 2006 — 40 с.
5. Рябинин И.А. Надежность и безопасность сложных систем. // СПб.: Политехника, 2000. – 248 с.
6. Bruce E.Briley, "A New Approach to Switching Network Synthesis," International Conference on Communications Technology, Shanghai, PRC, 1994.
7. Майника Э. Алгоритмы оптимизации на путях и графах. – М.:Мир, 1981. – 323 с.