УДК 519.217

DOI: 10.21779/2542-0321-2020-35-3-53-62

В.М. Захаров, С.В. Шалагин, Б.Ф. Эминов

Метод представления множеств стохастических матриц для многопараметрического анализа укрупненных и расширенных цепей Маркова

ФГБОУ ВПО «Казанский национальный исследовательский технический университет им. А.Н. Туполева – КАИ»; Россия, 420111, г. Казань, ул. Карла Маркса, 10; gilvv@mail.ru; sshalagin@mail.ru; bulfami@mail.ru

Решена задача представления множеств стохастических матриц укрупненных и расширенных цепей Маркова для кластерного анализа с применением программных средств системы STATISTICA. Задача кластерного анализа рассматривается как задача многопараметрической классификации множеств стохастических укрупненных и расширенных матриц по заданным параметрам их схожести или различия на основе методов прикладной многомерной математической статистики. Параметры – классификационные признаки – определяются как функционалы от элементов матриц и отражают структурные и асимптотические свойства стохастических матриц. Решением задачи является получение разбиения заданного множества стохастических матриц на группу кластеров. Предложен двухэтапный метод формирования различных множеств стохастических укрупненных и расширенных матриц, основанный на применении на первом этапе формирования в качестве модели автономного вероятностного автомата и на втором этапе – специальных разработанных алгоритмов построения стохастических укрупненных и расширенных матриц. Вероятностный автомат позволяет формировать разнообразие множеств стохастических матриц за счет изменения функции переходов и закона входной случайной величины. Приведена оценка мощности множества эргодических стохастических матриц, представляемых рассматриваемой автоматной моделью. Представлена методика многопараметрической классификации рассматриваемых стохастических матриц, включающая предложенные алгоритмы вычисления классификационных признаков.

Ключевые слова: *цепи Маркова, матрицы, многопараметрический анализ.*

Введение

Задача автоматизированного анализа многомерных данных актуальна при моделировании сложных систем (вычислительных и информационных систем, систем защиты информации, марковских систем массового обслуживания и др. [1–6]). Для описания подобных систем широко используются автоматные вероятностные модели (АВМ) [1, 4–6], задаваемые стохастическими матрицами. В связи с этим актуальной задачей анализа АВМ является задача классификации стохастических матриц по некоторым критериям (параметрам) их схожести или различия – задача кластерного анализа стохастических матриц, решение которой дает разбиение заданного множества стохастических матриц на группу кластеров [7–10]. Подход решения этой задачи, рассматриваемый в [7–10], основан на методах прикладной многомерной математической статистики [11–13] с применением программных средств вида «Интегрированная система STATISTICA» [13]. Для реализации этого подхода необходимо формировать различные заданные множества стохастических матриц (стохастические матрицы, описывающие регулярные, эргодические цепи Маркова [14], стохастические матрицы с различной энтропией [15] и др. [7–11]). Востребованной задачей в этом направлении является задача построения АВМ для моделирования укрупненных и расширенных цепей Маркова (ЦМ) [14, 16–18]. Представление стохастических процессов в виде укрупненных ЦМ позволяет изучать сложный процесс методами цепей Маркова [14] и решать различные прикладные задачи [3, 19–21]. В [16] предложены алгоритмы для автоматизированного построения укрупненных ЦМ на основе заданных регулярных стохастических матриц [14].

В [14] показано, что на заданной ЦМ можно сконструировать некоторую большую (расширенную) цепь, которая дает более детальную информацию о рассматриваемом первоначальном процессе. У расширенных ЦМ текущее состояние определяет поведение системы в будущем через конечное число шагов и может быть использовано для предсказания поведения системы. Автоматизированные алгоритмы построения расширенных цепей Маркова предложены в [18].

В [8–10] предложены алгоритмы построения эргодических стохастических матриц с заданными структурами и асимптотическими свойствами на основе автоматных вероятностных моделей. Однако вопросы по разработке алгоритмов построения требуемого разнообразия множеств стохастических матриц для многопараметрического анализа укрупненных и расширенных ЦМ, вопросы классификации данных стохастических матриц автоматизированными методами прикладной многомерной математической статистики исследованы недостаточно.

Целью работы является разработка метода формирования множеств стохастических матриц для многопараметрического анализа укрупненных и расширенных цепей Маркова с применением программных средств вида «Интегрированная система STATISTICA».

Построение множеств стохастических матриц укрупненных цепей Маркова

Определение укрупненной ЦМ. Пусть задана регулярная конечная ЦМ [14] системой

$$\left(S, P, \overline{\pi_0}\right), \tag{1}$$

где $S=\{s_i\}$ – конечное множество состояний ЦМ, $|S|=n$; $P=\left(p_{ij}\right)$, $i,j=\overline{0,n-1}$ – регулярная стохастическая матрица размера $n\times n$; p_{ij} – переходные вероятности ЦМ; $\overline{\pi_0}$ – вектор начального распределения вероятностей состояний ЦМ.

Выполним разбиение исходного множества состояний S ЦМ (1) на t непересекающиеся подмножества (на классы) вида $A=\{A_0, A_1, ..., A_{t-1}\}$, для которых

$$\bigcup_{j=0}^{t-1} A_j = S\,,\ A_j \cap A_d = 0\,, \tag{2}$$

где $\forall j, \forall d = \overline{0,t-1}$ и $j \neq d$.

Пусть каждое из подмножеств A_j, $j=\overline{0,t-1}$ будет новым состоянием укрупненной ЦМ (по терминологии [14]), а стохастический закон укрупненной ЦМ будет задаваться стохастической укрупненной матрицей $\hat{P}=\left(\hat{p}_{dj}\right)$ размера $t\times t$, $j,d=\overline{0,t-1}$. Обозначим множество состояний $Y=\{y_0, y_1, ..., y_{t-1}\}$ укрупненной ЦМ.

Пусть $p_{kA_j}=\sum\limits_{s_i\in A_j} p_{ki}$, $i,k=\overline{0,t-1}$, $j=\overline{0,t-1}$ – вероятность попасть из состояния s_i в множество A_j за один шаг исходной ЦМ (1).

В соответствии с работой [14] свойства регулярной стохастической матрицы, наличие которых при заданном разбиении множества состояний на непересекающиеся классы интерпретируется как возможность укрупнения стохастической матрицы, т. е. возможность укрупнения ЦМ.

Теорема 1 [14]. Для того чтобы состояния ЦМ можно было укрупнить посредством разбиения на классы $A=\left\{A_0, A_1, ..., A_{t-1}\right\}$, необходимо и достаточно, чтобы для любых двух классов A_d и A_j и для $\forall s_k \in A_d$, $k=\overline{0, n-1}$, $d, j=\overline{0,t-1}$, вероятности p_{kA_j} имели одно и то же значение. Вероятности $\left\{\hat{p}_{dj}\right\}$ переходов между классами образуют стохастическую матрицу $\hat{P}$ укрупненной ЦМ.

Определение 1. Цепь Маркова с регулярной стохастической матрицей *P*, удовлетворяющую условие теоремы 1, будем называть *укрупняемой*.

Определение 2. Цепь Маркова со стохастической матрицей $\hat{P}$, полученную укрупнением цепи Маркова (1) по разбиению (2), будем называть *укрупненной*.

Построение матрицы $\hat{P}$ можно провести в соответствии с теоретическим обоснованием из [14] по алгоритму [16].

Отметим следующие свойства укрупняемых цепей Маркова.

1) Если исходная регулярная ЦМ укрупняема по алгоритму [16], то результатом ее укрупнения также будет регулярная ЦМ [22].

2) Пусть ЦМ (1), заданная матрицей *P*, укрупняема по заданному разбиению (2). Получена матрица P^k возведением матрицы *P* в степень *k*, $2 \le k \le \infty$. Тогда ЦМ, заданная матрицей P^k, также является укрупняемой по разбиению (2) [22].

3) Стохастический предельный вектор $\overline{\pi_{np}}=\left(\pi_0^{np}, \quad \pi_1^{np}, \quad ..., \quad \pi_{n-1}^{np}\right)$ укрупняемой ЦМ и стохастический предельный вектор $\overline{\hat{\pi}_{np}}=\left(\hat{\pi}_0^{np}, \quad \hat{\pi}_1^{np}, \quad ..., \quad \hat{\pi}_{t-1}^{np}\right)$ укрупненной ЦМ связаны соотношением [22]

$$\overline{\hat{\pi}_{np}}=\overline{\pi_{np}}\cdot V, \tag{3}$$

где $V=\left(v_{ij}\right)$, $i=\overline{0, n-1}$, $j=\overline{0, t-1}$ – булева матрица размера $n\times t$, единичное значение элемента v_{ij} которого определяет, что состояние s_i исходной ЦМ входит в укрупненное состояние A_j: $v_{ij}=\begin{cases}1, s_i \in A_j \\ 0, s_i \notin A_j\end{cases}$.

Построение множеств стохастических матриц укрупняемых и укрупненных цепей Маркова

Рассмотрим следующий автономный вероятностный автомат [23].

$$ABA=\left(S, \hat{X}, \Delta(x,s)=s, \overline{\pi_0}\right), \tag{4}$$

где $S=\left\{s_i\right\}$, $i=\overline{0, n-1}$, – множество состояний; $\hat{X}$ – дискретная случайная величина

$\hat{X} = \begin{pmatrix} x_0 & x_1 & \dots & x_l \\ p_0 & p_1 & \dots & p_l \end{pmatrix}$, принимающая конечное множество значений $X = \{x_0 \quad x_1 \quad \dots \quad x_l\}$ на входе АВА (4) с вероятностями в виде вектора $\overline{P} = \{p_0 \quad p_1 \quad \dots \quad p_l\}$, $0 < p_i < 1$, $\sum_{i=0}^{l-1} p_i = 1$; $\Delta(x,s)$ – функция переходов АВА (1); $\overline{\pi_0}$ – вектор начального распределения вероятностей состояний ЦМ). $\Delta(x,s)$ задается автоматной таблицей размера $n \times l$ произвольно и по алгоритму [24] на основе разложения заданной регулярной стохастической матрицы $P = (p_{ij})$, $i, j = \overline{0, n-1}$, размера $n \times n$.

Последовательность состояний автомата (4) является цепью Маркова [24]. Матрицу *P* представим как разложение вида [24]

$$P_S = \sum_{k=0}^{l-1} p_k M(x_k), \tag{5}$$

где p_k, $k = \overline{0, l-1}$ – элементы вектора $\overline{P}$, $M(x_k)$, $k = \overline{0, l-1}$ – простая матрица размера $n \times n$, соответствующая символу x_k; *l* удовлетворяет соотношению [24]

$$l \le n^2 - n + 1. \tag{6}$$

Замечание. Оценка мощности множества автоматных таблиц размера $n \times l$ для задания функций $\Delta(x,s)$ определена на основе функции, определяющей число перестановок множества *S* с повторениями [25]:

$$C_K(i_1, \dots, i_n) = L_1 = \frac{K!}{i_1! \cdot i_2! \cdot \dots \cdot i_n!}, \tag{7}$$

где $K = n \cdot l$, i_1, i_2, …, i_n – натуральные числа, такие, что $\sum_{j=1}^{n} i_j = n \cdot l$.

Закон (стохастическую матрицу *P* размера $n \times n$) цепи Маркова, получаемый в автомате (4), можно однозначно вычислить в виде (5) в соответствии с [1] по заданным элементам автомата (4): $\hat{X}$ и $\Delta(x,s)$.

Алгоритм вычисления стохастической матрицы вида (5) (далее алгоритм 1) представим на основе соотношений (5) и (6) по заданной паре элементов $(\hat{X}, \Delta(x,s))$ следующими тремя этапами.

1. Определить по заданной $\Delta(x,s)$ множество простых матриц $M(x_k)$, $k = \overline{0, l-1}$.
2. Перемножить полученные $M(x_k)$ на соответствующие элементы p_k, $k = \overline{0, l-1}$ заданного вектора $\overline{P}$.
3. Вычислить P_S в соответствии с (5).

Задавая для (4) различные $\Delta(x,s)$ в виде соответствующих автоматных таблиц при фиксированном стохастическом векторе $\overline{P}$, можно получить по алгоритму 1 множество регулярных стохастических матриц *P*.

Меняя в (4) стохастический вектор $\overline{P}$ при фиксированной $\Delta(x,s)$, можно получить множество стохастических матриц *P* из фиксированного подкласса, определяемого функцией $\Delta(x,s)$.

Оценим общее число возможного разнообразия построенных стохастических матриц *P* на основе модели (4) по алгоритму 1. Введем ограничения. Пусть для вектора вида $\overline{P}$ элементы вычисляются согласно формуле:

$$p_i = a_i / N \, , \; \sum_{i=0}^{l-1} a_i = N \, , \qquad (8)$$

где a_i – целые неотрицательные числа и *N* кратно *l*. Тогда $\overline{P}$ относится к классу распределений вероятностей, мощность которого оценивается величиной [23].

$$L_2 = C_{l+N-1}^{N-1} \, . \qquad (9)$$

Величина (9) определяет мощность множества стохастических матриц, получаемых изменением вектора $\overline{P}$. Решение задачи построения множества рассматриваемых векторов с мощностью (9) можно выполнить в соответствии с [23]. Из соотношений (7) – (9) следует справедливость утверждения

Утверждение. Если мощность множества автоматных таблиц для задания $\Delta(x,s)$ в автомате (4) определяется величиной (7), то мощность множества стохастических матриц *P*, представляемых автоматной моделью (4) и удовлетворяющих условиям (8), (9), оценивается произведением $L_3 = L_1 \cdot L_2$.

Построение множества стохастических матриц *P* на модели (4) является первым этапом построения множества стохастических матриц $\hat{P}$.

Задача укрупнения ЦМ – задача второго этапа – связана с поиском подходящего разбиения в пространстве возможных разбиений множества *A* для фиксированной матрицы *P*. Проверку выполнения условия возможности укрупнения для стохастической матрицы *P* и решение задачи укрупнения ЦМ можно проводить по алгоритму [16].

Величину L_3 можно рассматривать как возможное пространство стохастических матриц *P*, на основе которых выполняется поиск укрупняемых ЦМ. Возможное пространство классов разбиения вида (2) для фиксированной матрицы *P* можно оценить по методике [16].

Построение множеств стохастических матриц расширенных цепей Маркова

Определение и представление расширенной цепи Маркова. Пусть задана последовательность состояний s_{j1}, s_{j2}, … простой однородной ЦМ с конечным множеством состояний $S = \{s_j\}$ и регулярной стохастической матрицей $P = (p_{ij})$ [14] размера $n \times n$, $i, j = \overline{0, n-1}$.

Рассматриваемые в этом разделе результаты могут быть перенесены в соответствии с [18] и на цепи Маркова, задаваемые эргодическими стохастическими матрицами. Образуем из этой исходной цепи по аналогии с [14] новую, расширенную цепь Маркова (РЦМ) следующим образом [18]. Составим всевозможные цепочки символов $s_j \in S$ длины $r = v + k$, $r \geq 2$, $v \geq 1$, $k \geq 1$, $j = \overline{0, n-1}$, имеющих вид (s_{j1} , s_{j2}, …, s_{jv}, $s_{j(v+1)}$, …, $s_{j(v+k)}$). Смежные цепочки содержат $k = r - v$ общих символов и *v* («величина сдвига») отличающихся символов.

Предположим, что исходная цепь за $2v + k - 1$ шагов последовательно переходит из некоторого состояния s_{j1} в s_{j2}, затем из s_{j2} в s_{j3}, …, из $s_{j(2v+k-1)}$ в $s_{j(2v+k)}$.

Смежные цепочки длины r будем рассматривать как один шаг перехода нового процесса из состояния (s_{j1} , s_{j2}, …, s_{jv}, $s_{j(v+1)}$, …, $s_{j(v+k)}$) в состояние ($s_{j(v+1)}$, …, $s_{j(2v)}$, $s_{j(2v+1)}$, …, $s_{j(2v+k)}$). Этот новый расширенный процесс (расширенная цепь Маркова) является ЦМ с n^{v+k} состояниями (цепочки символов s_j длины r, $j=\overline{0,n-1}$). Обозначим состояния расширенной цепи Маркова символами y_i, из алфавита $Y=\{y_i\}$, $i=\overline{0,n^{v+k}-1}$.

Обозначим через Q стохастическую расширенную матрицу размера $n^{v+k}\times n^{v+k}$ этой РЦМ. Определим матрицу Q расширенной цепи Маркова через матрицу P исходной ЦМ при $v\geq 1$, $k\geq 1$ в соответствии с [18]. Для частного случая, $v=k=1$, эта задача решена в [14].

Получение матрицы Q расширенной цепи Маркова осуществим на основе понятия «промежуточной» стохастической матрицы [18] размера $n^{v+k}\times n^{v+k}$, вычисляемой на основе регулярной матрицы P исходной ЦМ. Промежуточную матрицу обозначим как $W=\left(w_{ij}\right)$, $i,j=\overline{0,n^r-1}$, $r\geq 2$.

Ненулевые элементы матрицы W вычисляются по формуле [18]:

$$w_{i,(i\cdot n \bmod n^r+d)}=p_{(i \bmod n),d}, \tag{10}$$

где i – текущий номер строки матрицы W, $i,j=\overline{0,n^r-1}$, d – текущий номер столбца матрицы P, $d=\overline{0,n-1}$, $r\geq 2$.

Матрица Q расширенной цепи Маркова связана с матрицей W соотношением [18]

$$Q=W^v, \tag{11}$$

где W^v – v-тая степень матрицы W, $v\geq 1$, $k\geq 1$.

Пример. По заданной матрице $P=\begin{pmatrix}3/8 & 5/8\\ 6/8 & 2/8\end{pmatrix}$ и заданным $r=3$ и $v=2$ можно построить на основе соотношения (10) следующую матрицу Q расширенной цепи Маркова:

$$Q=\begin{pmatrix}
9/64 & 15/64 & 15/32 & 5/32 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 9/32 & 15/32 & 3/16 & 1/16\\
9/64 & 15/64 & 15/32 & 5/32 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 9/32 & 15/32 & 3/16 & 1/16\\
9/64 & 15/64 & 15/32 & 5/32 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 9/32 & 15/32 & 3/16 & 1/16\\
9/64 & 15/64 & 15/32 & 5/32 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 9/32 & 15/32 & 3/16 & 1/16
\end{pmatrix}$$

Отметим следующие свойства расширенной цепи Маркова, представляемой матрицей (11) [18].

Теорема 2 [18]. Пусть дана матрица $P=\left(p_{ij}\right)$, $\forall p_{ij}>0$, $i,j=\overline{0,n-1}$ и на ее основе получена по алгоритму (10) матрица Q расширенной цепи Маркова размера $n^r\times n^r$, $r=v+k\geq 2$, $v\geq 1$, $k\geq 1$. Тогда данная расширенная цепь Маркова регулярная.

Предельный стохастический вектор матрицы Q расширенной цепи Маркова определяется по аналогии с [14, 183] соотношением

$$\overline{\pi_{np}} = \overline{\pi_{np}^{(w)}} \cdot W\,, \tag{12}$$

где $\overline{\pi_{np}^{(w)}}$ – предельный вектор матрицы *W*.

Построение множества расширенных стохастических матриц

Решение задачи построения множества стохастических матриц Q расширенных цепей Маркова будем рассматривать в виде метода, включающего следующие два этапа.

1) Формируется множество эргодических стохастических матриц $P=\left(p_{ij}\right)$, $\forall p_{ij}>0$, $i,j=\overline{0,n-1}$ на модели (4) с применением алгоритма 1. Мощность получаемого множества стохастических матриц *P* оценим произведением $L_3 = L_1 \cdot L_2$.

2) Каждой матрице *P* ставится в соответствие по формулам (10), (11) матрица вида (11) при заданных $r = v + k \geq 2$, $v \geq 1$, $k \geq 1$.

Методика кластерного анализа стохастических матриц

Решение задачи многопараметрической классификации множества стохастических матриц укрупненных или расширенных цепей Маркова методами кластерного анализа можно представить по методике, включающей три следующих этапа.

Этап 1. Определение множества классификационных признаков и алгоритмов для вычисления параметров (значений признаков) стохастических матриц для кластерного анализа.

Этап 2. Классификация множеств стохастических матриц методами кластерного анализа [7, 13]: определение метрических функций, определение методов кластерного анализа, получение кластерного решения в виде разбиения заданного множества стохастических матриц на группу кластеров.

Этап 3. Оценка качества кластерного решения – оценка однородности анализируемого множества стохастических матриц по некоторым критериям (параметрам) схожести или различия их структур.

На этапе 1 определяемые признаки представим в виде трех типов, которые являются функционалами от положительных элементов стохастических матриц $\hat{P}$ или Q и отображают [8]:

1) уровень, характеризуемый мерой отклонения положительных элементов укрупненных матриц $\hat{P}$ или расширенных матриц Q от нуля, – тип $C_1 = n \max\limits_{d,j=\overline{0,n-1}} \hat{p}_{dj}$ (для матриц $\hat{P}$) и $C_1 = n \max\limits_{i,j=\overline{0,n-1}} w_{ij}$ (для матриц Q);

2) рассеяние (разброс) элементов в строках и столбцах элементов матриц $\hat{P}$ или Q относительно средних – тип $C_2 = \sum_{j=0}^{n-1}\sum_{d=0}^{n-1}\left(\hat{p}_{dj} - \overline{\hat{p}_j}\right)^2$, где $\overline{\hat{p}_j} = (n)^{-1}\sum_{d=0}^{n-1}\hat{p}_{dj}$ (для матриц $\hat{P}$) и $C_2 = \sum_{j=0}^{n-1}\sum_{d=0}^{n-1}\left(w_{dj} - \overline{w_j}\right)^2$, где $\overline{w_j} = (n)^{-1}\sum_{d=0}^{n-1}w_{dj}$ (для матриц Q);

3) энтропию матриц [15] $\hat{P}$ и матриц Q – тип $C_3 = -\sum_{d=0}^{n-1}\sum_{j=0}^{n-1}\alpha_d \hat{p}_{dj}\log_2\left(\hat{p}_{dj}\right)$, где элементы α_d предельного вектора вычисляются по соотношению (3) для матрицы $\hat{P}$ и $C_3 = -\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}\alpha_i w_{ij}\log_2\left(w_{ij}\right)$, где элементы α_i предельного вектора вычисляются по соотношению (12) для матрицы Q.

На этапе 2 постановка задачи кластерного анализа в общем виде заключается в том, чтобы на основании данных, вычисленных по формулам C_1, C_2, C_3 для соответствующего построенного множества стохастических матриц $\hat{P}$ или Q и представленных в виде таблицы «объект-признак» [7, 13], и одной из метрик [13] получить дивизивным методом «*K*-средних» [7; 13] с применением программных средств вида «Интегрированная система STATISTICA» [13] разбиение анализируемого множества стохастических матриц на непересекающиеся подмножества.

На этапе 3 оценка качества кластерного решения определяется на основе критерия вида [13] с применением программных средств вида «Интегрированная система STATISTICA»:

$$F = \left(d_b/n_b\right)/\left(d_w/N\right),$$

где d_b, d_w – суммы межкластерных и внутрикластерных расстояний, $n_b = k(k-1)/2$ – количество межкластерных расстояний. Критерий показывает, во сколько раз среднее межкластерное расстояние больше среднего внутрикластерного расстояния.

Заключение

Рассмотрена задача кластерного анализа укрупненных и расширенных цепей Маркова как задача многопараметрической классификации укрупненных и расширенных стохастических соответствующих матриц по некоторым параметрам схожести или их различия на основе методов прикладной многомерной математической статистики с применением программных средств системы STATISTICA. Решением задачи кластерного анализа является получение разбиения заданного множества стохастических матриц на группу кластеров. Необходимые для классификации множества стохастических матриц, описывающих укрупненные или расширенные цепи Маркова, формируются в два этапа. На первом этапе формируется множество регулярных стохастических матриц на автономном вероятностном автомате по предложенному алгоритму. Представленная автоматная модель позволяет формировать разнообразие множеств стохастических матриц за счет изменения функции переходов и случайного входа в вероятностном автомате. Приведена оценка мощности множества стохастических матриц, представляемых рассматриваемой автоматной моделью.

На втором этапе на основе полученного на первом этапе множества стохастических матриц формируются по заданным алгоритмам множества стохастических укрупненных или расширенных матриц. Разработана методика многопараметрической классификации стохастических матриц с применением программных средств системы STATISTICA, включающая алгоритмическое представление параметров – классификационных признаков, описывающих стохастические матрицы.

Работа выполнена при финансовой поддержке гранта РФФИ N 18-01-00120 «Специализированные устройства для генерирования и обработки массивов данных в архитектуре программируемых логических интегральных схем класса FPGA».

Литература

1. *Левин Б.Р., Шварц В.* Вероятностные модели и методы в системах связи и управления. – М.: Радио и связь, 1985. – 312 с.

2. *Герасименко В.А.* Защита информации в автоматизированных системах обработки данных. Кн. 1. – М.: Энергоатомиздат, 1994. – 400 с.

3. *Погорелов Б.А., Пудовкина М.А.* Об обобщениях марковского подхода при изучении алгоритмов блочного шифрования // Прикладная дискретная математика. Приложение. – 2014. – № 7. – С. 51–52.

4. *Бухараев Р.Г.* Основы теории вероятностных автоматов. – М.: Наука, 1985. – 287 с.

5. *Феррари Д.* Оценка производительности вычислительных систем. – М.: Мир, 1981. – 576 с.

6. *Глова В.И., Захаров В.М., Песошин В.А., Шалагин С.В.* Моделирование. Вероятностные дискретные модели: уч. пособие. – Казань: Абак, 1998. – 52 с.

7. *Барковский С.С., Захаров В.М., Лукашов А.М., Нурутдинова А.Р., Шалагин С.В.* Многомерный анализ данных методами прикладной статистики: учеб. пособие. – Казань: Издательство КГТУ, 2010. – 122 с.

8. *Захаров В.М., Нурмеев Н.Н., Салимов Ф.И., Шалагин С.В.* Классификация стохастических эргодических матриц методами кластерного и дискриминантного анализа // Исследования по информатике. – 2000. – Вып. 2. – С. 91–106.

9. *Нурутдинова А.Р., Шалагин С.В.* Многопараметрическая классификация автоматных марковских моделей на основе генерируемых ими последовательностей состояний // Прикладная дискретная математика. – 2010. – № 4. – С. 41–54.

10. *Шалагин С.В., Нурутдинова А.Р.* Сравнительный анализ вычислительной сложности идентификации конечных простых однородных цепей Маркова // Вестник технологического университета. – 2016. – Т. 19, № 13. – С. 153–156.

11. *Дюран Б., Оделл П.* Кластерный анализ. – М.: Статистика, 1977. – 128 с.

12. *Енюков И.С.* Методы, алгоритмы, программы многомерного статистического анализа: Пакет ППСА. – М.: Финансы и статистика, 1986. – 230 с.

13. *Боровиков В.П.* Statistica: искусство анализа данных на компьютере – 2-е изд. – СПб.: Питер, 2003. – 700 с.

14. *Кемени Дж., Снелл Дж.,* Конечные цепи Маркова. – М.: Наука, 1970. – 272 с.

15. *Хинчин А.Я.* Понятие энтропии в теории вероятностей // Успехи математических наук. – 1953. – № 3 (55). – С. 3–20.

16. *Захаров В.М., Эминов Б.Ф.* Алгоритмы укрупнения цепей Маркова // Вестник Казанского государственного технического университета им. А.Н. Туполева, – 2013 – № 2, вып. 1, – С.125– 133.

17. *Zakharov V.M., Shalagin S.V., Eminov B.F.* Representing Lumped Markov Chains by Minimal Polynomials over Field GF(q) // Journal of Physics: Conference Series. – 2018 – V. 1015 – P. 032033. – DOI: https://doi.org/10.1088/1742-6596/1015/3/032033.

18. *Эминов Б.Ф., Захаров В.М.* Полиномиальные модели расширенных цепей Маркова над полем GF(2^n) // Системы управления и информационные технологии. – 2014. – № 2 (56). – С. 26–30.

19. *Geiger B., Temmel C.* Lumpings of Markov chains, entropy rate preservation, and higher-order lumpability // Advances in Applied Probability. – 2014. – № 51 (4). – P. 1114–1132.

20. *Максимов Ю.И.* Некоторые результаты для задачи укрупнения состояний цепей Маркова // Труды по дискретной математике. – 2004. – № 8. – С. 148–154.

21. *Katehakis M., Smit L.* A Successive Lumping Procedure for a Class of Markov Chains // Probability in the Engineering and Informational Sciences – 2012. – № 26 (4), – P. 483–508.

22. *Эминов Б.Ф., Захаров В.М.* Об асимптотических свойствах укрупняемых и укрупненных цепей Маркова // Вестник технологического университета. – Т. 18, № 10. – 2015. – С. 167–173.

23. *Бухараев Р.Г., Захаров В.М.* Управляемые генераторы случайных кодов. – Казань: Издательство КГУ, 1978.

24. *Ченцов В.М.* Об одном методе синтеза автономного стохастического автомата // Кибернетика. – 1968. – № 3. – С. 32–35.

25. *Бронштейн И.Н, Семендяев К.А.* Справочник по математике для инженеров и учащихся втузов. – 13-е изд., испр. – М.: Наука; Гл. ред. физ.-мат. лит., 1986. – 544 с.

Поступила в редакцию 18 марта 2020 г.

UDC 519.217

DOI: 10.21779/2542-0321-2020-35-3-53-62

Method for Submitting Sets of Stochastic Matrices for Multi-parametric Analysis of Aggregated and Extended Markov's Chains

V.M. Zakharov, S.V. Shalagin, B.F. Eminov

Kazan National Research Technical University named after A.N.Tupolev – KAI; Russia, 420111, Kazan, Karl Marks st., 10; gilvv@mail.ru; sshalagin@mail.ru; bulfami@mail.ru

The problem of representing a set of stochastic matrices aggregated and extended Markov chains to cluster analysis software system STATISTICA is solved in the article. The problem of cluster analysis is considered as the problem of multiparametric classification of sets of stochastic aggregated and extended matrices according to the specified parameters of their similarity or difference, based on the methods of applied multidimensional mathematical statistics. Parameters-classification features are defined as functionals from matrix elements and reflect the structural and asymptotic properties of stochastic matrices. The solution of the problem is to obtain a partition of a given set of stochastic matrices into a group of clusters. The two-step method of forming different sets of stochastic aggregated and expanded matrices based on the application of the first stage of formation as a model of autonomous probabilistic automaton and the second stage are special developed algorithms for constructing stochastic aggregated and extended matrices is proposed. The probabilistic automaton allows forming a variety of sets of stochastic matrices by changing the transition function and the law of the input random variable. The power estimation of the set of ergodic stochastic matrices represented by the considered automaton model is given. The method of multiparametric classification of the considered stochastic matrices including the proposed algorithms for calculating the classification features is presented.

Keywords: *Markov`s chains, matrices, multiparametric analysis.*

Received 18 March 2020