

Learned and Self-Learning Recognition, Management and Decision-Making Systems on Base of the Logical Neural Networks

It is examined the logic neural networks application for associative "non-formula" hardly formulated tasks and simulation computing, in medical question-answered system, in user identification systems, in the step by step adaptive routing for wireless telecommunications and transport networks, in self-learning control systems by fuzzy data. In the base of logical neural networks lies the mathematical logic of events, combining the initial information from mathematical logic and probability theory. Excitation of receptors is based on the reliability of statements about data ownership, which provides a single system of measurement of baseline data and dramatically extends the scope of application of neural networks to construct almost any decision-making system that combines the data of any nature, types and forms of representation. Excitation of neurons in the output layer indicates a decision to be made in the form of value, the text of the control signal, and so on. Networks are reduced to a single layer. In these learning situations directly linked to corresponding solutions. Logic neural networks are differing by low requirements for the qualification developer, simple calculations, high speed and the possibility of parallelization. Guidelines for a better incarnation of some of the known decision-making systems are provides.

Keywords: logic neural network, the activation function, the matrix sequence, associative computing, information and referral system, self-learning management system

References

1. Barskij A. B. *Nejronnye seti: raspoznavanie, upravlenie, prinyatie reshenij*. M.: Finansy i statistika, 2004. 175 p.
2. Barskij A. B. *Logicheskie nejronnye seti*. M.: INTUIT; BINOM. Laboratoriya znaniy, 2007. 352 p.
3. Barskij A. B. *Matematicheskaja logika sobytij i logicheskie nejronnye seti*. *Informacionnye tehnologii. Prilozhenie*, 2007, no. 7. 32 p.
4. Barskij A. B. *Nejronnye seti logicheskogo vyvoda*. Kurs lekczij. Saarbrücken: LAP LAMBERT Academic Publishing, 2011. 400 p.
5. Pospelov D. A. *Modelirovanie racyzhdenij. Opyt analiza myslitelnyh aktov*. M.: Radio i svyaz', 1989. 184 p.
6. Pospelov D. A. *Situacionnoe upravlenie. Teoriya i praktika*. M.: Nauka. Gl. red. Fiz.-mat. Lit, 1986. 288 p.
7. Barskij A. B., Nguen Van Loj. *Informacionno-spravochnaja sistema na logicheskoj nejronnoj seti "Mnogoservernaya baza dannyh s czirkulirujushhimi segmentami"*. *Informacionnye tehnologii*. 2013, no. 9. P. 57–61.
8. Barskij A. B., Dmitriev A. A., Barskaya O. A. *Medicinskie informacionno-spravochnye sistemy na logicheskikh nejronnyh setyah*. *Informacionnye tehnologii. Prilozhenie*, 2010, no. 1. 32 p.
9. Chzho Vin Latt. *Logicheskaya nejronnaya set' v osnove identifikacii pol'zovatelya v sisteme "klient — server" na baze tryoh harakteristik "pocherka"*. *Estestvennye i tehicheskie nauki*, 2010, no. 3.
10. Barskij A. B., Said Mohammed M. N. *Vybor smeshheniya pri poshagovoj marshrutizacii v besprovodnoj seti*. *Mir Transporta*, 2013, no. 2, pp. 30–37.
11. Zamyshlyayev A. M. *Prikladnye informacionnye sistemy upravleniya nadyozhnost'yu, bezopasnost'yu, riskami i resursami na zhelezno-dorozhnom transporte*. M.: "Zhurnal "Nadyozhnost'", 2013. 143 p.
12. Barskij A. B. *Rejtingovye sistemy na logicheskikh nejronnyh setyah*. *Informacionnye tehnologii*. 2014, no. 11, pp. 66–72.

УДК 004.942

Л. С. Куравский, д-р техн. наук, проф., декан, e-mail: l.s.kuravsky@gmail.com,
П. А. Мармалюк, канд. техн. наук, доц., Г. А. Юрьев, канд. физ.-мат. наук, доц., П. Н. Думин, аспирант,
Московский городской психолого-педагогический университет,
факультет информационных технологий

Методы численной идентификации марковских моделей и их сравнительный анализ¹

Представлены численные методы нулевого и первого порядка, предназначенные для идентификации марковских процессов с дискретными состояниями и непрерывным временем по результатам наблюдений, а также технология вычислительного эксперимента, обеспечивающего сравнение характеристик эффективности рассматриваемых подходов. Анализ результатов проведенных экспериментов показал, что разработанные методы идентификации имеют преимущества перед классическим градиентным методом первого порядка. Рассмотренные алгоритмы можно применять для настройки параметров обучаемых структур, включая нейронные сети.

Ключевые слова: марковские модели, идентификация моделей, многомерная нелинейная оптимизация, обучаемые структуры

¹ Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект № 14-06-00191) и Российского гуманитарного научного фонда (проект № 14-06-12012).

1. Постановка задачи идентификации

Поведение многих технических, диагностических и других систем представляется параметрическими математическими моделями в форме систем обыкновенных дифференциальных уравнений. В частности, такая форма представления применялась для моделей усталостного разрушения и тестирования, построенных на базе *марковских случайных процессов с дискретными состояниями и непрерывным временем* [1–6, 11]. Важной особенностью практического использования марковских моделей в указанных задачах является то, что их параметры идентифицируются по результатам наблюдений, представленных наборами значений исследуемых функций в контрольных точках.

Марковские модели для описания динамики переходов между состояниями представляются ориентированными графами, в которых вершины соответствуют состояниям, а дуги — переходам, для которых выполняются свойства пуассоновских потоков событий. В этих потоках число событий N , попадающих в любой временной интервал длины τ , начинающийся в момент t , распределено согласно закону Пуассона:

$$P_{t,\tau}(N = r) = \frac{a(t, \tau)^r}{r!} e^{-a(t, \tau)},$$

где $P_{t,\tau}(X = r)$ — вероятность появления r событий в течение рассматриваемого интервала; $a(t, \tau)$ — среднее число событий, попадающих в интервал длины τ , начинающийся в момент времени t . Рассматриваются только стационарные потоки, в которых $a(t, \tau) = \eta\tau$, а $\eta = \text{const}$ есть интенсивность стационарного потока. Упомянутые выше предположения о свойствах потоков событий обычны для прикладных задач [7], так как эти потоки (или потоки, близкие к ним по свойствам) часто встречаются на практике благодаря предельным теоремам для потоков событий.

Полагается, что:

- для указанных процессов с $n + 1$ дискретными состояниями и непрерывным временем заданы начальные распределения вероятностей и наблюдаемые частоты пребывания в состояниях процессов $\{F_{id}\}_{i=0, \dots, n}$ в моменты времени $\{t_d\}_{d=0, \dots, D-1}$, где D — количество моментов времени, в которые фиксировались частоты F_{id} ; $0 \leq t_d \leq T$, T — конечный момент времени;

- интенсивности переходов между состояниями полностью или частично являются неизвестными (свободными) параметрами.

Динамика изменения вероятностей пребывания в состояниях процесса определяется системой обыкновенных дифференциальных уравнений Колмогорова в матричной форме:

$$\frac{dp(t)}{dt} = \mathbf{M}(\lambda)\mathbf{p},$$

где $0 \leq t \leq T$, $\mathbf{p}(t) = (p_0(t), \dots, p_n(t))^T$ — вероятности пребывания в состояниях процесса; $\lambda = (\lambda_0, \dots, \lambda_m)^T$ — упорядоченный набор интенсивностей переходов между состояниями, $n > m$; \mathbf{M} — матрица интенсивностей переходов между состояниями порядка $n + 1$.

Для указанной системы уравнений ставится задача идентификации набора параметров λ . Значения параметров из этого набора определяются путем сравнения наблюдаемых и прогнозируемых гистограмм, описывающих распределения частот пребывания в состояниях модели, а именно: вычисляются значения, обеспечивающие наилучшее соответствие наблюдаемых и ожидаемых частот попадания в определенное состояние системы в заданные моменты времени. При этом определяется набор интенсивностей λ , обеспечивающий наименьшее значение статистики Пирсона:

$$X^2(\lambda) = \sum_{d=0}^{D-1} \sum_{i=0}^n \frac{(p_i(\lambda t_d)N - F_{id})^2}{p_i(\lambda t_d)N},$$

где $N = \sum_{i=0}^n F_{id}$. Эта статистика используется как

мера соответствия модели наблюдениям.

Доказано, что при выполнении ряда общих условий значения указанной статистики X^2 , получаемые при подстановке истинных решений, асимптотически описываются распределением χ^2 с $n - m$ степенями свободы, где m — число определяемых параметров, причем вычисленные значения свободных параметров при увеличении объема выборки сходятся по вероятности к искомому решению [8]. Это позволяет использовать статистику Пирсона для проверки гипотезы о том, что полученный прогноз согласуется с результатами наблюдений.

Таким образом, постановка задачи предполагает решение обратной задачи, в которой коэффициенты дифференциальных уравнений определяются по фрагментам решения, заданным наблюдениями. Представленную выше постановку можно формально рассматривать как прямую задачу с усложненным вычислением критерия X^2 посредством решения задачи Коши для системы обыкновенных дифференциальных уравнений, которая решается стандартными методами оптимизации.

Для ее решения, в частности, допустимы алгоритмы локальной оптимизации, включая градиентные методы, и стохастические алгоритмы. Однако опыт их практического применения позволяет говорить о неприемлемо большом времени вычислений в случае задач высокой размерности (при больших значениях параметра m), а также о проблемах, обусловленных возможной неединственностью получаемого решения.

Для решения этих проблем предложены представленные далее в подразделах 2.2 и 2.3 численные методы [10], предназначенные для идентификации набора параметров λ , определяющего матрицу \mathbf{M} .

Особенностью этих методов является использование начальных приближений, вычисляемых по результатам наблюдений, а также оценок чувствительности критерия оптимизации к малым изменениям идентифицируемых параметров. Новым элементом второго из предложенных подходов является улучшение временных характеристик применяемого комбинаторного алгоритма подбора значений идентифицируемых переменных за счет замены на каждой итерации вычислений перебора различных сочетаний этих значений в окрестности их текущих оценок перебором значений только тех переменных, к изменению которых наиболее чувствителен критерий оптимизации. Для оценки этой чувствительности используют разностные аппроксимации соответствующих частных производных. Метод, рассмотренный в подразделе 2.2, относится к методам нулевого порядка, а его модифицированный вариант из подраздела 2.3 — к методам первого порядка. Характеристики их эффективности сопоставляют с аналогичными показателями классического подхода, представленного градиентным методом первого порядка из подраздела 2.1.

Рассмотренные алгоритмы можно эффективно применять для настройки параметров обучаемых структур, включая нейронные сети.

2. Сравнимые методы оптимизации

2.1. Градиентный метод: алгоритм вычислений

1. Используя имеющиеся результаты наблюдений, вычислить в качестве начальных оценок параметров λ_i ($i = 0, \dots, m$), обозначаемых как λ_i^0 ($i = 0, \dots, m$), выборочные средние значения числа переходов между парами соответствующих состояний марковского процесса в единицу времени (или, что иногда удобнее, величины, обратные выборочным средним продолжительностям времени между двумя смежными переходами для этих же пар состояний). Полученные оценки рассматривать как начальные приближения к идентифицируемым параметрам.

2. Задать счетчик итераций $j = 0$.

3. Задать начальный шаг градиентного метода h .

4. В точке текущего приближения λ^j вычислить оценку градиента $\nabla X^2(\lambda)$ с использованием центрально-разностной аппроксимации, при этом значения $\mathbf{p}(\lambda, t_{ij})$ определить путем численного решения

задачи Коши для системы уравнений $\frac{d\mathbf{p}}{dt} = \mathbf{M}(\lambda)\mathbf{p}$

с заданными начальными условиями.

5. Выполнить коррекцию текущего приближения:

$$\lambda^{j+1} = \lambda^j - h\nabla X^2(\lambda^j).$$

6. Если $X^2(\lambda^{j+1}) < X^2(\lambda^j)$, $\|\lambda^j - \lambda^{j+1}\| > \varepsilon$ и не исчерпан лимит итераций, то положить $j = j + 1$ и перейти к шагу 3, иначе перейти к шагу 6.

7. Если $X^2(\lambda^{j+1}) \geq X^2(\lambda^j)$, $\|\lambda^j - \lambda^{j+1}\| > \varepsilon$ и не исчерпан лимит итераций, то положить $h = \frac{h}{2}$ и перейти к шагу 4, иначе завершить вычисления.

2.2. Метод полного перебора: алгоритм вычислений

1. Используя имеющиеся результаты наблюдений, вычислить в качестве начальных оценок параметров λ_i ($i = 0, \dots, m$), обозначаемых как λ_i^0 ($i = 0, \dots, m$), выборочные средние значения числа переходов между парами соответствующих состояний марковского процесса в единицу времени (или, что иногда удобнее, величины, обратные выборочным средним продолжительностям времени между двумя смежными переходами для этих же пар состояний). Полученные оценки рассматривать как начальные приближения к идентифицируемым параметрам.

2. Положить $j = 1$, $k = 1$, $\beta_1 = 1 + \gamma_1$, где γ_1 — параметр алгоритма.

3. Для каждой из текущих оценок идентифицируемых параметров λ_i^j ($i = 0, \dots, m$) на j -й итерации алгоритма вычислить верхнюю λ_i^{j+} и нижнюю λ_i^{j-}

границу сдвига оценки по формулам $\lambda_i^{j+} = \beta_j \lambda_i^j$,

$\lambda_i^{j-} = \beta_j^{-1} \lambda_i^j$.

4. Выполнив полный перебор всех вариантов текущих оценок набора параметров $\lambda_j = (\lambda_0^j, \dots, \lambda_m^j)^T$,

в которых каждый из его компонентов λ_i^j ($i = 0, \dots, m$) принимает только три возможных значения из множества $\{\lambda_i^{j-1,-}, \lambda_i^{j-1}, \lambda_i^{j-1,+}\}$, выбрать из указанных

вариантов одну из оценок $\lambda_{j,*} = (\lambda_{0,*}^j, \dots, \lambda_{m,*}^j)^T$, обеспечивающих минимальное значение критерия X^2 , равное $X_{j,\min}^2$.

5. Если для всех компонентов выбранного набора $\lambda_{j,*}$ выполняется равенство $\lambda_{i,*}^j = \lambda_i^{j-1}$ ($i = 0, \dots, m$), то положить $k = k + 1$, $\gamma_k = \gamma_{k-1}/2$.

6. Положить $j = j + 1$, $\beta_j = 1 + \gamma_k$, $\lambda_j = \lambda_{j-1,*}$.

7. Если $X_{j-1,\min}^2 > X_*^2$, где X_*^2 — параметр алгоритма, то перейти к шагу 3, иначе завершить вычисления.

Временная сложность одной итерации рассмотренного алгоритма есть $O(3^{m+1})$, что приводит к неприемлемо длительным вычислениям при больших значениях m . Для того чтобы временная сложность стала приемлемой для практических вычислений, перебор вариантов текущих оценок набора

параметров $\lambda_j = (\lambda_0^j, \dots, \lambda_m^j)^T$ на шаге 4 приведенного выше алгоритма следует осуществлять не для всех параметров λ_i^j ($i = 0, \dots, m$), а только для их подмножества, состоящего из $l < m + 1$ элементов, к изменению которых наиболее чувствителен критерий X^2 .

2.3. Метод перебора значимых параметров: алгоритм вычислений

1. Используя имеющиеся результаты наблюдений, вычислить в качестве начальных оценок параметров λ_i ($i = 0, \dots, m$), обозначаемых как λ_i^0 ($i = 0, \dots, m$), выборочные средние значения числа переходов между парами соответствующих состояний марковского процесса в единицу времени (или, что иногда удобней, величины, обратные выборочным средним продолжительностям времени между двумя смежными переходами для этих же пар состояний). Полученные оценки рассматривать как начальные приближения к идентифицируемым параметрам.

2. Положить $j = 1, k = 1, \beta_1 = 1 + \gamma_1$, где γ_1 — параметр алгоритма.

3. Оценить чувствительность критерия X^2 в малой окрестности каждой из текущих оценок идентифицируемых параметров λ_i^j ($i = 0, \dots, m$) на j -й итерации алгоритма, используя для этого разностные аппроксимации абсолютных значений частных производных $\delta_i(\varepsilon) = |X^2(\lambda_0^j, \dots, \lambda_i^j + \varepsilon, \dots, \lambda_m^j) - X^2(\lambda_0^j, \dots, \lambda_i^j, \dots, \lambda_m^j)|/\varepsilon$, где $i = 0, \dots, m$; ε — параметр алгоритма.

4. Выбрать l текущих оценок идентифицируемых параметров $\lambda_{i_q}^j$ (где $i_q \in \{0, \dots, m\}; q = 1, \dots, l$; l — параметр алгоритма); имеющих наибольшие значения оценок чувствительности $\delta_i(\varepsilon)$.

5. Для каждой из выбранных на шаге 4 текущих оценок идентифицируемых параметров $\lambda_{i_q}^j$ на j -й итерации алгоритма, где $i_q \in \{0, \dots, m\}; q = 1, \dots, l$, вычислить верхнюю $\lambda_{i_q}^{j+}$ и нижнюю $\lambda_{i_q}^{j-}$ границу сдвига оценки по формулам $\lambda_{i_q}^{j+} = \beta_j \lambda_{i_q}^j, \lambda_{i_q}^{j-} = \beta_j^{-1} \lambda_{i_q}^j$.

6. Выполнив полный перебор всех вариантов текущих оценок набора параметров $\lambda_j = (\lambda_0^j, \dots, \lambda_m^j)^T$, в которых каждый из его компонентов $\lambda_{i_q}^j$, где $i_q \in \{0, \dots, m\}; q = 1, \dots, l$, принимает только три возможных значения из множества $\{\lambda_{i_q}^{j-1,-}, \lambda_{i_q}^{j-1}, \lambda_{i_q}^{j-1,+}\}$, а оставшиеся компоненты принимают значение

своей текущей оценки на j -й итерации алгоритма, выбрать из указанных вариантов одну из оценок $\lambda_{j,*} = (\lambda_{0,*}^j, \dots, \lambda_{m,*}^j)^T$, обеспечивающих минимальное значение критерия X^2 , равное $X_{j,\min}^2$.

7. Если для всех компонентов выбранного набора $\lambda_{j,*}$ выполняется равенство $\lambda_{i,*}^j = \lambda_{i,*}^{j-1}$ ($i = 0, \dots, m$), то положить $k = k + 1, \gamma_k = \gamma_{k-1}/2$.

8. Положить $j = j + 1, \beta_j = 1 + \gamma_k, \lambda_j = \lambda_{j-1,*}$.

9. Если $X_{j-1,\min}^2 > X_*^2$, где X_*^2 — параметр алгоритма, то перейти к шагу 3, иначе завершить вычисления.

Для решаемых на практике типовых задач общее время вычислений зависит от параметра l монотонно, достигая наименьшего значения при $l = 1$. Временная сложность одной итерации модифицированного алгоритма при $l = 1$ есть $O(m)$, что свидетельствует об асимптотически линейной зависимости времени вычислений, необходимого для выполнения одной итерации, от числа идентифицируемых параметров.

3. Генерация пула данных для вычислительного эксперимента

Для оценки эффективности предложенных алгоритмических решений был создан пул тестовых структур, близких по своим параметрам и структуре к моделям, используемым при решении реальных прикладных задач. В качестве базовой рассматривалась марковская модель, организованная по схеме "гибели и размножения". В процессе моделирования управляемыми параметрами были число состояний такой модели и число независимых параметров, через которые выражались интенсивности переходов между состояниями.

Самым простым случаем является ситуация, при которой все интенсивности λ равны между собой, а самым сложным — случай, при котором каждый из переходов имеет собственную интенсивность. В процессе генерации моделей из заданного числового интервала случайным образом выбирались значения переменных, соответствующие интенсивностям переходов между состояниями.

Пример типовой модели представлен на рис. 1 в виде ориентированного графа.

Структура связей подобного графа может быть описана матрицей смежности, элементы которой содержат значения интенсивностей переходов (рис. 2).

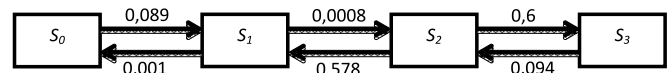


Рис. 1. Организованная по схеме "гибели и размножения" марковская модель из четырех состояний. Интенсивности переходов определяются шестью независимыми параметрами

	i	S_0	S_1	S_2	S_3
j	S_0		0,089		
	S_1	0,001		0,0008	
	S_2		0,578		0,6
	S_3			0,094	

Рис. 2. Пример матрицы смежности, отражающей структуру связей и интенсивности переходов, для модели с четырьмя состояниями и шестью независимыми параметрами

Матрица, представленная на рис. 2, позволяет сформировать следующую систему обыкновенных дифференциальных уравнений Колмогорова:

$$\frac{dP_{S_0}(t)}{dt} = P_{S_1}(t) \cdot 0,001 - P_{S_0}(t) \cdot 0,089;$$

$$\frac{dP_{S_1}(t)}{dt} = P_{S_0}(t) \cdot 0,089 + P_{S_2}(t) \cdot 0,578 - P_{S_1}(t) \cdot (0,001 + 0,0008);$$

$$\frac{dP_{S_2}(t)}{dt} = P_{S_2}(t) \cdot 0,0008 + P_{S_3}(t) \cdot 0,094 - P_{S_2}(t) \cdot (0,578 + 0,6);$$

$$\frac{dP_{S_3}(t)}{dt} = P_{S_2}(t) \cdot 0,6 - P_{S_3}(t) \cdot 0,094.$$

Для всех созданных моделей в начальный момент времени t_0 вероятность пребывания в состоянии S_0 была равна единице, а во всех прочих — нулю. Для каждой возможной комбинации "Число состояний/число параметров" (до 10 состояний) было сгенерировано по 20 примеров (рис. 3).

Для систем уравнений Колмогорова, соответствующих сгенерированным моделям, на временном интервале $[0; 1000]$ численно решалась задача

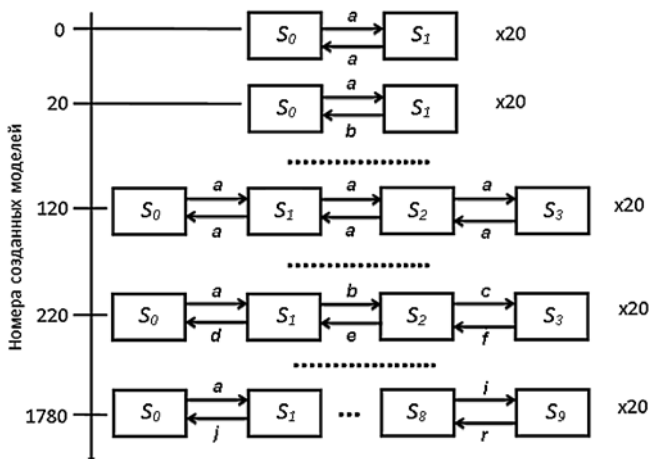


Рис. 3. Структура сформированного пула моделей для идентификации

Коши [9]. Вычисленные значения вероятностей $\mathbf{P}(t) = (p_{S_0}(t), \dots, p_{S_n}(t))^T$ для контрольных точек $\{t_d\}_{d=0, \dots, D-1}$ сохранялись в файле модели.

Следует отметить, что в ряде комбинаций подобный подход к генерации моделей приводит к нарушению указанного в разд. 1 условия $n > m$. Использование статистики χ^2 в качестве критерия качества модели при этом становится невозможным. В таких случаях рекомендуется использовать рассмотренный в работе [1] критерий, основанный на оценке значимости компонентов уравнения регрессии, связывающего наблюдаемые и прогнозируемые частоты.

Чтобы избежать ряда вычислительных проблем, связанных с представлением малых вероятностей, вместо χ^2 в качестве минимизируемого критерия использовали статистику $Q = \sum_{d=0}^{D-1} \sum_{i=0}^n (p_i(t_d, \lambda) - \tilde{p}_i(t_d))^2$,

где $\tilde{p}_i(t_d)$ — вероятность пребывания в i -м состоянии модели в момент времени t_d , полученная при интегрировании уравнений модели на этапе ее генерации.

Для вычислительного эксперимента было сгенерировано 1800 моделей, каждая из которых представлялась XML-файлом, содержащим: матрицу смежности; значения независимых параметров, определяющих интенсивности переходов; параметры, используемые при численном интегрировании уравнений модели; имена состояний; значения вероятностей $\mathbf{P}(t) = (p_{S_0}(t), \dots, p_{S_n}(t))^T$ в контрольных точках $\{t_d\}_{d=0, \dots, D-1}$.

Таблицы 1 и 2 содержат число сгенерированных моделей для каждого набора параметров и каждого числа состояний.

Таблица 1		Таблица 2	
Число моделей, сгенерированных для каждого числа параметров		Число моделей, сгенерированных для каждого числа состояний	
Число параметров	Число моделей	Число состояний	Число моделей
1	180	2	40
2	180	3	80
3	160	4	120
4	160	5	160
5	140	6	200
6	140	7	240
7	120	8	280
8	120	9	320
9	100	10	360
10	100		
11	80		
12	80		
13	60		
14	60		
15	40		
16	40		
17	20		
18	20		

4. Параметры вычислительной системы

Программное обеспечение для вычислительного эксперимента реализовано на языке *G* в среде графического программирования *National Instruments LabVIEW 2010*. Характеристики системы, работавшей под управлением ОС *Microsoft Windows 7 Professional (SP1)*, следующие:

- материнская плата ASRock Z77 Extreme4-M Socket LGA 1155 (1066—2800 МГц);
- четырехъядерный процессор Intel Core i5-3330 с частотой 3000 МГц;
- два модуля оперативной памяти DDR3 Corsair XMS3 объемом 4 Гбайт с частотой 1333 МГц.

5. Параметры процедур идентификации

Поскольку при генерации пула данных для вычислительного эксперимента отсутствует информация, необходимая для определения начальных оценок интенсивностей переходов, оценки параметров λ_i^0 на первом шаге представленных выше алгоритмов устанавливались равными заданным константам. Значение параметра γ_1 , определяющего сдвиг от текущих оценок идентифицируемых величин, устанавливалось равным 0,9. Число контрольных моментов времени D было равно 100. По достижении значения Q , удовлетворяющего неравенству

$$\frac{Q}{(n+1)D} \leq 0,001, \text{ поиск останавливался, а соответствующая модель считалась идентифицированной.}$$

Время вычислений, затрачиваемое на каждую модель, ограничивалось 200 с. По его истечении поиск останавливался, а соответствующая модель считалась неидентифицированной. В случаях методов полного перебора и перебора двух значимых параметров идентификацию проводили только для моделей, имеющих не более шести параметров.

6. Результаты вычислительного эксперимента

После идентификации каждой модели характеристики проведенного процесса вычислений сохранялись в виде *XML*-файла. Сформированный таким образом набор данных включал название метода идентификации, число состояний, число параметров модели, временные отметки для каждой итерации и другую необходимую информацию.

Созданные файлы использовали для формирования выборки результатов вычислительных экспериментов со следующими характеристиками: число состояний *States* и параметров *Params* модели, идентификационный номер модели *Trial*, евклидово расстояние *Distance* начальных оценок параметров до их значений, заложенных при генерации модели, число итераций *Iterations*, выполненных до достижения одного из критериев останова, результирующее значение *ResFit* минимизируемого критерия,

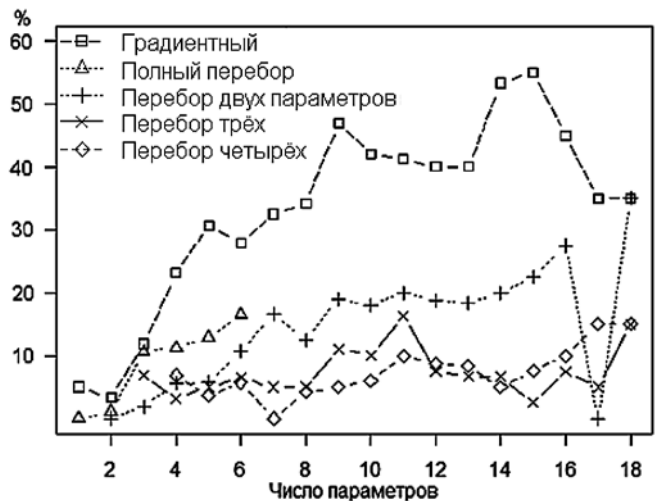


Рис. 4. Зависимости доли неидентифицированных моделей от числа параметров моделей при разных методах идентификации

целевое значение *GoalFit* минимизируемого критерия, булев флаг *MissFitFlag*, определяющий успешность идентификации модели (0 — модель успешно идентифицирована, 1 — модель не идентифицирована), номинальный фактор *Alg* с пятью значениями (0 — градиентный метод, 1 — полный перебор, 2 — перебор двух значимых параметров, 3 — перебор трёх значимых параметров, 4 — перебор четырёх значимых параметров). Полученные данные анализировали в программной среде для статистических вычислений *R* [12].

Зависимости доли моделей, которые не были идентифицированы вследствие превышения заданного лимита времени, от числа параметров моделей при разных методах идентификации представлены на рис. 4. Эти результаты свидетельствуют о том, что разработанные методы идентификации, за исключением метода полного перебора при более чем пяти параметрах, обеспечивают существенное ускорение вычислений по сравнению с рассмотренным классическим градиентным методом.

На рис. 5 представлены графики типа "ящик с усами", отражающие зависимости статистических характеристик времени вычислений (размаха, межквартильного интервала и медианы) от числа состояний и используемого метода идентификации (слева) и от числа параметров модели и используемого метода идентификации (справа). Вдоль горизонтальной оси на этом и последующем рисунках указаны составные показатели, первый компонент которых есть число состояний или параметров, а второй — индекс метода (1 — полный перебор, 2 — перебор двух значимых параметров). Приведенные графики иллюстрируют показательный характер зависимости времени вычислений от числа параметров модели для метода идентификации, основанного на полном переборе комбинаций оценок,

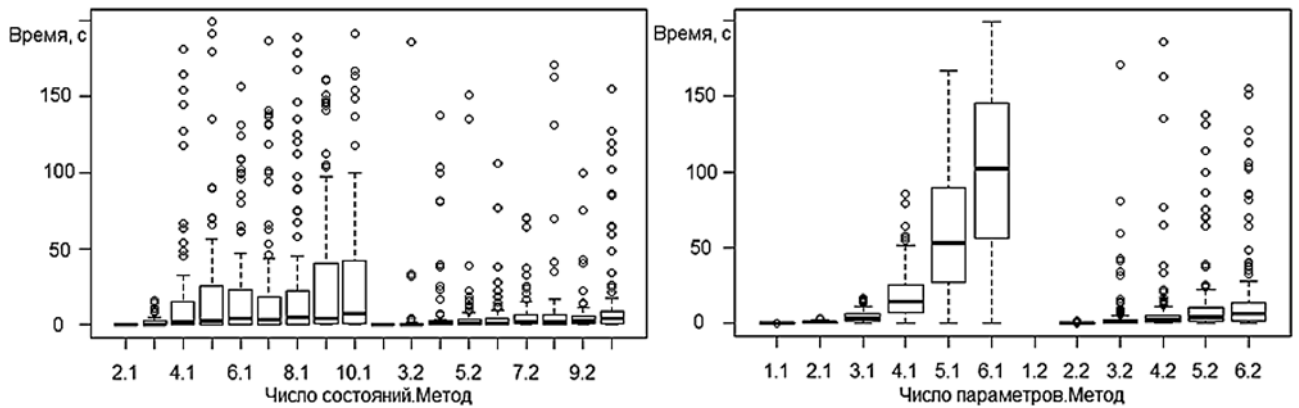


Рис. 5. Зависимости статистических характеристик времени вычислений (размаха, межквартильного интервала и медианы) от числа состояний и параметров и используемого метода идентификации (1 — полный перебор; 2 — перебор двух значимых параметров)

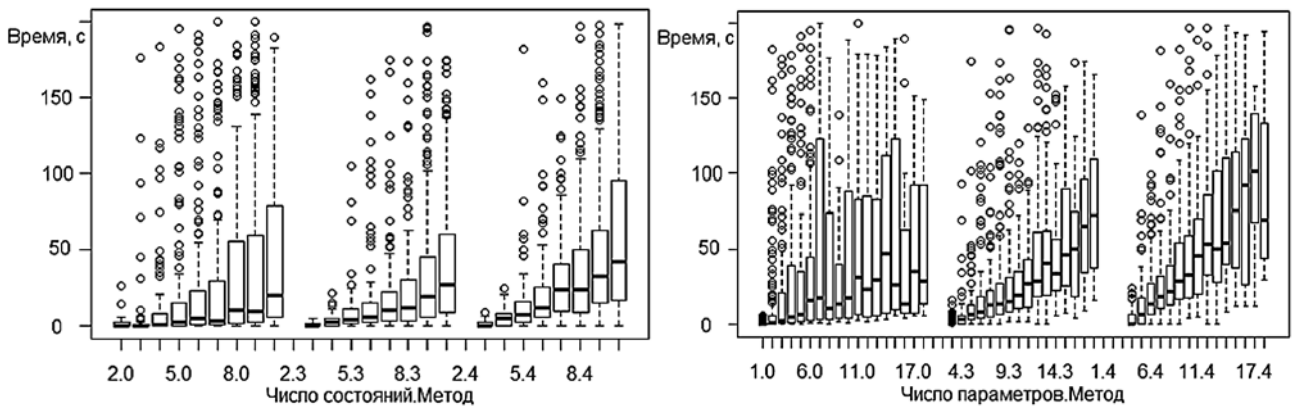


Рис. 6. Зависимости статистических характеристик времени вычислений (размаха, межквартильного интервала и медианы) от числа состояний и параметров и используемого метода идентификации (0 — градиентный метод, 3 — метод перебора трех значимых параметров, 4 — метод перебора четырех значимых параметров)

и близкий к линейному характер такой зависимости в случае перебора двух значимых параметров.

Те же самые статистические характеристики для градиентного метода и методов перебора трех и четырех значимых параметров, обозначенных, соответственно, индексами 0, 3 и 4, представлены на рис. 6. Динамика роста медианы времени вычислений на приведенных графиках позволяет говорить о том, что градиентный метод наименее подвержен влиянию числа идентифицируемых параметров. Однако зависимости, представленные на рис. 4, свидетельствуют, что этот метод в несколько раз чаще, чем другие подходы, не обеспечивает нахождение решения (если число параметров больше восьми, решение не определяется примерно в 34...55 % случаев). Это существенно ограничивает практическую применимость градиентного метода и, по-видимому, объясняется известной чувствительностью этого способа решения к выбору начальных приближений.

Для удобства сопоставления рассмотренных методов идентификации на рис. 7 отдельно приведены зависимости среднего времени вычислений от числа параметров моделей. Представленные результаты свидетельствуют о том, что разработанные методы

идентификации имеют преимущества перед классическим градиентным методом при числе идентифицируемых параметров, не превышающем 15. Если число указанных параметров превышает 15, то градиентным методом решение получается не-

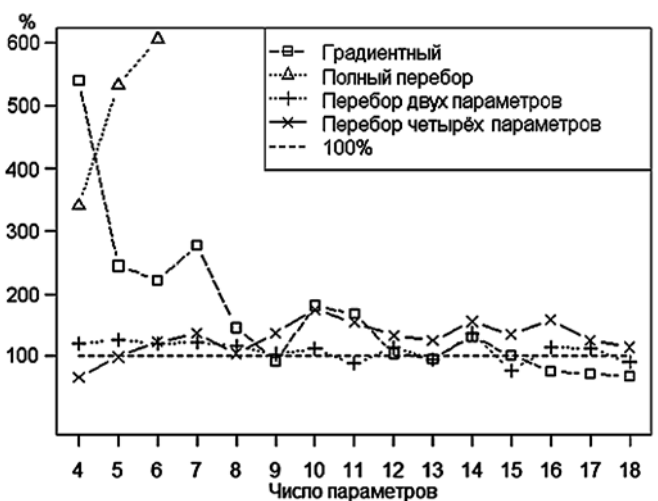


Рис. 7. Зависимости доли среднего времени вычислений относительно среднего времени, затраченного на вычисления методом перебора трех параметров, от числа параметров моделей для различных методов идентификации

Таблица 3

Статистические оценки значимости влияний различных характеристик на время вычислений, полученные путем анализа заданной линейной регрессионной зависимости

Статистика $R^2 = 0,26$; F -статистика = 309,3 при 7 и 6026 степенях свободы, $p < 2,2 \cdot 10^{-16}$					
Параметр регрессионной модели	Оценка коэффициента	Стандартная ошибка	t -статистика	p -значение	Статистическая значимость
b_0	-5,94	1,88	-3,15	0,00165	Есть
<i>States</i>	1,65	0,24	6,86	$7,95 \cdot 10^{-12}$	Есть
<i>Params</i>	3,95	0,14	27,27	$< 2,0 \cdot 10^{-16}$	Есть
<i>Distance</i>	14,35	1,34	10,7	$< 2,0 \cdot 10^{-16}$	Есть
<i>Alg1</i>	1,52	1,65	0,91	0,36	Нет
<i>Alg2</i>	-17,55	1,45	-12,1	$< 2,0 \cdot 10^{-16}$	Есть
<i>Alg3</i>	-21,43	1,5	-14,29	$< 2,0 \cdot 10^{-16}$	Есть
<i>Alg4</i>	-12,71	1,56	-8,16	$4,21 \cdot 10^{-16}$	Есть

сколько быстрее, однако примерно в 35—55 % случаев его сходимость не гарантирована (см. рис. 4), и решение за разумное время не удастся получить.

В табл. 3 представлены статистические оценки значимости влияний различных характеристик на время вычислений, полученные для следующей линейной регрессионной зависимости:

$$Time = b_0 + b_1 States + b_2 Params + b_3 Distance + b_4 Alg1 + b_5 Alg2 + b_6 Alg3 + b_7 Alg4,$$

где *Time* — время вычислений; *States* — число состояний модели; *Params* — число параметров модели; *Distance* — евклидово расстояние от начальных оценок параметров до их оптимальных значений; *Alg1*, *Alg2*, *Alg3*, *Alg4* — бинарные показатели, принимающие значение 1 для результатов, полученных, соответственно, полным перебором, перебором двух значимых параметров, перебором трех значимых параметров и перебором четырех значимых параметров, и принимающие значение 0 — в противном случае.

Данные, представленные в табл. 3, свидетельствуют о том, что рассмотренная регрессионная модель в целом, и все ее компоненты статистически значимы. Анализ коэффициентов, соответствующих бинарным показателям, представляющим метод идентификации, позволяет говорить о существенном ускорении вычислений при использовании методов перебора значимых параметров по сравнению с классическим градиентным методом.

В целом рассмотренные выше результаты вычислительных экспериментов позволяют говорить о заметных преимуществах разработанных методов идентификации перед классическим градиентным методом.

7. Основные результаты и выводы

1. Разработаны численные методы нулевого и первого порядка, предназначенные для идентификации марковских процессов с дискретными состояниями и непрерывным временем по результатам наблюдений.

2. Разработана технология вычислительного эксперимента в целях сравнения характеристик эффективности рассматриваемых подходов к численному решению задачи идентификации. Ее программная реализация включает средства для генерации пула марковских моделей, исследования их поведения и анализа получаемых результатов.

3. Анализ результатов проведенных вычислительных экспериментов показал, что разработанные методы идентификации имеют преимущества перед классическим градиентным методом первого порядка.

Список литературы

1. Куравский Л. С., Марголис А. А., Юрьев Г. А., Мармалюк П. А. Концепция системы поддержки принятия решений для психологического тестирования // Психологическая наука и образование. 2012. № 1. С. 56—65.
2. Куравский Л. С., Мармалюк П. А., Алхимов В. И., Юрьев Г. А. Математические основы нового подхода к построению процедур тестирования // Экспериментальная психология. 2012. Том 5, № 4. С. 75—98.
3. Куравский Л. С., Марголис А. А., Мармалюк П. А., Юрьев Г. А., Думин П. Н. Обучаемые марковские модели в задачах оптимизации порядка предъявления психологических тестов // Нейрокомпьютеры: разработка и применение. 2013. № 4. С. 28—38.
4. Куравский Л. С., Мармалюк П. А., Алхимов В. И., Юрьев Г. А. Новый подход к построению интеллектуальных и компетентностных тестов // Моделирование и анализ данных. 2013. № 1. С. 4—28.
5. Куравский Л. С., Мармалюк П. А., Барабанщиков В. А., Безруких М. М., Демидов А. А., Иванов В. В., Юрьев Г. А. Оценка степени сформированности навыков и компетенций на основе вероятностных распределений глазодвигательной активности // Вопросы психологии. 2013. № 5. С. 64—81.
6. Марковские модели в задачах диагностики и прогнозирования: учеб. пособие / Под ред. Л. С. Куравского. М.: РУСАВИА, 2013. 172 с.
7. Овчаров Л. А. Прикладные задачи теории массового обслуживания. М.: Машиностроение, 1969. 324 с.
8. Крамер Г. Математические методы статистики. М.: Мир, 1976. 648 с.
9. Формалев В. Ф., Ревизников Д. Л. Численные методы. М.: ФИЗМАТЛИТ, 2004. 400 с.
10. Kuravsky L. S., Marmalyuk P. A., Yuryev G. A., Dumina P. N. A Numerical Technique for the Identification of Discrete-State Continuous-Time Markov Models // Applied Mathematical Sciences. 2015. Vol. 9. N. 8. P. 379—391.
11. Kuravsky L. S., Marmalyuk P. A., Baranov S. N., Alkhimov V. I., Yuryev G. A., Artyukhina S. V. A New Technique for Testing Professional Skills and Competencies and Examples of its Practical Applications // Applied Mathematical Sciences. 2015. Vol. 9. N. 21. P. 1003—1026.
12. The R Project for Statistical Computing. URL: <http://r-project.org/> (дата обращения: 25.04.2014).

L. S. Kuravsky, Professor, P. A. Marmalyuk, Professor Associated,
G. A. Yuryev, Professor Associated, P. N. Dumin, Graduate
Moscow State University of Psychology and Education, Computer Science Faculty,
127051 Moscow, Russia l.s.kuravsky@gmail.com

Methods of Numerical Identification for Markov Models and Their Comparative Analysis¹

Zero- and first-order numerical methods intended for identification of Markov processes with discrete states and continuous time using observation data as well as the computer experiment technology providing comparison of performance characteristics of the methods in question are under consideration. The first-order methods developed are characterized by utilizing both initial approximations derived from observation data and estimates of minimized criterion sensitivity to small variations of identified parameters. Analysis of the computer experiments' results showed the advantages of the developed methods over the classical first order gradient descent method. Improved running time of the methods is achieved through replacing, at each iteration, of enumeration of various combinations of parameter values in the neighborhood of their current estimates by enumeration of values of only those parameters to which the minimized criterion is highly sensitive. The algorithms under consideration can be used for tuning parameters of the trained structures including neural networks.

Keywords: Markov models, models identification, multivariate non-linear optimization, trained structures

References

1. Kuravskii L. S., Margolis A. A., Yur'ev G. A., Marmalyuk P. A. Kontsepsiya sistemy podderzhki prinyatiya reshenii dlya psikhologicheskogo testirovaniya, *Psikhologicheskaya nauka i obrazovanie*, 2012, no. 1, pp. 56–65.
2. Kuravskii L. S., Marmalyuk P. A., Alkhimov V. I., Yur'ev G. A. Matematicheskie osnovy novogo podkhoda k postroeniyu protsedur testirovaniya, *Ekspierental'naya psikhologiya*, 2012. Vol. 5, no. 4, pp. 75–98.
3. Kuravskii L. S., Margolis A. A., Marmalyuk P. A., Yur'ev G. A., Dumin P. N. Obuchaemye markovskie modeli v zadachakh optimizatsii poryadka pred'yavleniya psikhologicheskikh testov, *Neirokomp'yutery: razrabotka i primeneniye*, 2013, no. 4, pp. 28–38.
4. Kuravskii L. S., Marmalyuk P. A., Alkhimov V. I., Yur'ev G. A. Novyi podkhod k postroeniyu intellektual'nykh i kompetentnostnykh testov, *Modelirovaniye i analiz dannykh*, 2013, no. 1, pp. 4–28.
5. Kuravskii L. S., Marmalyuk P. A., Barabanshchikov V. A., Bezrukikh M. M., Demidov A. A., Ivanov V. V., Yur'ev G. A. Ot-senka stepeni sformirovannosti navykov i kompetentsii na osnove veroyatnostnykh raspredelenii glazodvigatel'noi aktivnosti, *Voprosy psikhologii*, 2013, no. 5, pp. 64–81.
6. Markovskie modeli v zadachakh diagnostiki i prognozirovaniya: Ucheb. posobie. Ed. L. S. Kuravskogo. Moscow, RUSAVIA, 2013, 172 p.
7. Ovcharov L. A. *Prikladnye zadachi teorii massovogo obsluzhivaniya*, Moscow, Mashinostroeniye, 1969. 324 p.
8. Kramer G. *Matematicheskie metody statistiki*. M.: Mir, 1976. 648 p.
9. Formalev V. F., Reviznikov D. L. *Chislennyye metody*. M.: FIZ-MATLIT, 2004, 400 p.
10. Kuravsky L. S., Marmalyuk P. A., Yuryev G. A., Dumin P. N. A Numerical Technique for the Identification of Discrete-State Continuous-Time Markov Models. *Applied Mathematical Sciences*, 2015, vol. 9, no. 8, pp. 379–391.
11. Kuravsky L. S., Marmalyuk P. A., Baranov S. N., Alkhimov V. I., Yuryev G. A., Artyukhina S. V. A New Technique for Testing Professional Skills and Competencies and Examples of its Practical Applications, *Applied Mathematical Sciences*. 2015, vol. 9, no. 21, pp. 1003–1026.
12. The R Project for Statistical Computing. URL: <http://r-project.org/> (дата обращения: 25.04.2014).

¹ This work was supported by the Russian Foundation for Basic Research (project № 14-06-00191) and the Russian Foundation for Humanities (project № 14-06-12012).

Адрес редакции:

107076, Москва, Стромьинский пер., 4

Телефон редакции журнала (499) 269-5510

E-mail: it@novtex.ru

Технический редактор *Е. В. Конова*.

Корректор *Е. В. Комиссарова*.

Сдано в набор 06.08.2015. Подписано в печать 23.09.2015. Формат 60×88 1/8. Бумага офсетная.

Усл. печ. л. 8,86. Заказ ИТ1015. Цена договорная.

Журнал зарегистрирован в Министерстве Российской Федерации по делам печати, телерадиовещания и средств массовых коммуникаций.

Свидетельство о регистрации ПИ № 77-15565 от 02 июня 2003 г.

Оригинал-макет ООО "Авансед солюшнз". Отпечатано в ООО "Авансед солюшнз".

119071, г. Москва, Ленинский пр-т, д. 19, стр. 1.