УДК 681.518

В.Р. Милов , В.Г. Баранов, А.Ю. Эпштейн, И.В. Шалашов

ПРОГНОЗИРОВАНИЕ СОСТОЯНИЯ ДИСКРЕТНЫХ СТОХАСТИЧЕСКИХ СИСТЕМ В УСЛОВИЯХ НЕОПРЕДЕЛЕННОСТИ НА ОСНОВЕ БАЙЕСОВСКОЙ МЕТОДОЛОГИИ

Нижегородский государственный технический университет им. Р.Е. Алексеева

Рассматривается вероятностно-статистический подход к прогнозированию состояния технических систем. Разработан прототип программного обеспечения для прогнозирования состояния систем, описываемых скрытыми марковскими моделями.

Ключевые слова: динамическая байесовская сеть, скрытая марковская модель, фильтрация, прогнозирование, структурно-параметрическое обучение.

Модели дискретных стохастических систем

Для описания функционирования стохастических систем с памятью находят применение модели в пространстве переменных состояния. Один из развивающихся способов моделирования заключается в использовании динамических байесовских сетей (ДБС). Динамическая байесовская сеть может быть представлена как пара $\{B_1, B_2\}$, где B_1 — байесовская сеть, задающая априорное распределение $P(\mathbf{z}(1))$; B_2 — байесовская сеть, состоящая из двух слоев и определяющая вероятности переходов

$$P(\mathbf{z}(n) \mid \mathbf{z}(n-1)) = \prod_{i=1}^{R} P(z_i(n) \mid Pa(z_i(n))). \tag{1}$$

Здесь $\mathbf{z} = [z_1, \ldots, z_R]^{\mathrm{T}}$ — вектор случайных величин, используемых для описания стохастической системы; $Pa(z_i(n))$ — множество родительских переменных для i-й переменной.

Зачастую множество переменных Z может быть представлено как объединение множества непосредственно не наблюдаемых переменных состояния системы X и множества регистрируемых величин Y, характеризующих состояние. При этом для описания системы обычно задается модель состояния, описывающая динамику изменения во времени вероятностей нахождения системы в различных состояниях, и модель наблюдения, описывающая связь наблюдений с состоянием системы. Эти модели определяются условными распределениями вероятностей $P(\mathbf{x}(n+1) \mid \mathbf{x}(n))$ и $P(\mathbf{y}(n) \mid \mathbf{x}(n))$ соответственно, а также распределением вероятностей $P(\mathbf{x}(1))$ состояния в начальный момент времени.

На рис. 1 представлены примеры графических моделей, соответствующих ДБС с единственной переменной состояния и единственной переменной наблюдения: скрытая марковская модель (СММ) и авторегрессионная СММ.

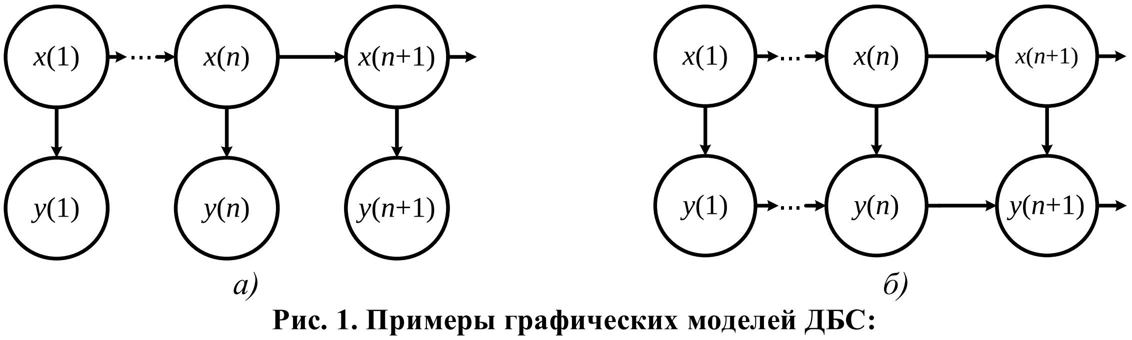


Рис. 1. Примеры графических моделей ДБС:
а – СММ; *б* – авторегрессионная СММ

Скрытая марковская модель соответствует ДБС с единственной дискретной переменной состояния, которая недоступна для наблюдения и определяются следующими распределениями и параметрами:

Распределение вероятностей начального состояния $\pi_i = P(x(1) = a_i)$.

Модель состояния, задаваемая с помощью $(r \times r)$-матрицы переходов $\mathbf{A}$ с элементами $A_{i,j} = P(x(n+1) = a_j \mid x(n) = a_i),\ i,\ j = \overline{1,\ r}$, где r – количество состояний.

Модель наблюдения в случае дискретных наблюдений задается с помощью $(r \times h)$-матрицы $\mathbf{B}$ с элементами $B_{i,j} = P(y(n) = b_j \mid x(n) = a_i),\ i = \overline{1,\ r},\ \ j = \overline{1,\ h}$, где h – количество возможных наблюдаемых значений. В случае действительных наблюдений часто используется гауссовское распределение с условной плотностью вероятности $p(\mathbf{y}(n) = \xi \mid x(n) = a_i) = N(\xi;\ \mu_i,\ \Sigma_i)$, где μ_i — математическое ожидание и $\mathbf{\Sigma}_i$ – ковариационная матрица для i-го состояния системы.

Процедуры прогнозирования и фильтрации

В многочисленных практически важных задачах интерес представляет прогнозирование состояния стохастической системы. Для этого находятся вероятности

$$P(x(n+\tau+1) \mid \mathbf{Y}_{1:n}) = \sum_{x(n+\tau)} P(x(n+\tau+1) \mid x(n+\tau)) P(x(n+\tau) \mid \mathbf{Y}_{1:n}),\ \ \tau = \overline{0,\ \tau_{\Pi} - 1}. \tag{2}$$

Здесь $\mathbf{Y}_{1:n}$ — последовательность наблюдений $y(1), \ldots, y(n)$, величина τ_{Π} определяет горизонт прогнозирования. Основу для прогнозирования представляет процедура фильтрации, в результате которой находятся вероятности

$$\begin{aligned} P(x(n) \mid \mathbf{Y}_{1:n}) &= \eta\, P(y(n) \mid x(n), \mathbf{Y}_{1:n-1}) P(x(n) \mid \mathbf{Y}_{1:n-1}) = \\ &= \eta\, P(y(n) \mid x(n)) \left[\sum_{x(n-1)} P(x(n) \mid x(n-1)) P(x(n-1) \mid \mathbf{Y}_{1:n-1}) \right], \end{aligned} \tag{3}$$

где $\eta = 1 / P(y(n) \mid \mathbf{Y}_{1:n-1})$ — константа нормализации.

Для случая СММ с дискретными переменными процедуры прогнозирования и фильтрации могут быть представлены в матричном виде

$$\mathbf{P}(x(n+\tau+1) \mid \mathbf{Y}_{1:n}) = \mathbf{A}^{\mathrm{T}}\, \mathbf{P}(x(n+\tau) \mid \mathbf{Y}_{1:n}),\ \ \tau = \overline{0,\ \tau_{\Pi} - 1}, \tag{4}$$

$$\mathbf{P}(x(n) \mid \mathbf{Y}_{1:n}) = \eta\, \mathbf{O}(n) \mathbf{A}^{\mathrm{T}}\, \mathbf{P}(x(n-1) \mid \mathbf{Y}_{1:n-1}),\ \ \mathbf{P}(x(1) \mid \mathbf{Y}_1) = \eta\, \mathbf{O}(1) \boldsymbol{\pi}. \tag{5}$$

Здесь $\mathbf{P}(x \mid \mathbf{Y}) = [P(x = a_1 \mid \mathbf{Y}),\ \ldots,\ P(x = a_r \mid \mathbf{Y})]^{\mathrm{T}}$, $\mathbf{O}(n) = \{O_{i,j}(n)\}$ — диагональная матрица с элементами $O_{i,i}(n) = B_{i,\tilde{j}}$, где $\tilde{j} = \arg\{y(n) = b_j\}$.

Структурно-параметрическое обучение

При наличии неопределенности модель системы определена не полностью. В общем случае возникает необходимость определения как параметров $\boldsymbol{\theta}$, так и структуры s модели [1]. На основе доступной обучающей выборки $\mathbf{D}$ может быть выполнена процедура обучения.

В условиях, когда обучающая выборка содержит значения не только переменных наблюдения $y(n)$, но и состояния $x(n)$, т.е. $\mathbf{D} = \{\mathbf{X},\ \mathbf{Y}\}$, оценка параметров при фиксированной структуре s может быть найдена по критерию минимума среднего риска, например, как апостериорное среднее, либо по критерию максимального правдоподобия

$$\hat{\boldsymbol{\theta}}_{ML} = \arg\max_{\boldsymbol{\theta}} L(\mathbf{D} \mid \boldsymbol{\theta},\ s). \tag{6}$$

Здесь $L(\mathbf{D} \mid \boldsymbol{\theta},\ s)$ — логарифм функции правдоподобия, определяемый выражением

$$L(\mathbf{D}|\boldsymbol{\theta}, s) \equiv \ln P(\mathbf{D}|\boldsymbol{\theta}, s) \equiv \ln P(\mathbf{X}, \mathbf{Y}|\boldsymbol{\theta}, s) =$$
$$= \ln P(x(1)|\boldsymbol{\theta}, s) + \sum_{n=2}^{N} \ln P(x(n)|x(n-1), \boldsymbol{\theta}, s) + \sum_{n=1}^{N} \ln P(y(n)|x(n), \boldsymbol{\theta}, s). \quad (7)$$

Задача определения структуры модели может быть решена по критерию максимума апостериорной вероятности. В распространенном случае различные структуры модели полагаются априори равновероятными. При этом оптимальная структура модели доставляет максимум логарифму маргинальной функции правдоподобия

$$L(\mathbf{D}|s) \equiv \ln P(\mathbf{D}|s) = \ln \int_{\boldsymbol{\Theta}} P(\mathbf{D}|\boldsymbol{\theta}, s) p(\boldsymbol{\theta}|s) d\boldsymbol{\theta}. \quad (8)$$

В общем случае интегрирование в (8) выполняется приближенно. Достаточно часто для выбора структуры модели находит применение информационный критерий Акаике

$$L_{AIC}(\mathbf{D}|s) = L(\mathbf{D}|\hat{\boldsymbol{\theta}}_{ML}, s) - \gamma. \quad (9)$$

При использовании байесовского информационного критерия целевая функция структурной оптимизации определяется выражением

$$L_{BIC}(\mathbf{D}|s) = L(\mathbf{D}|\hat{\boldsymbol{\theta}}_{ML}, s) - \frac{\gamma}{2} \ln N, \quad (10)$$

где N – объем обучающей выборки; γ – эффективное количество параметров. Например, при определении порядка q цепи Маркова эффективное количество параметров γ, согласно [2], составляет

$$\gamma = r^{q}(r-1). \quad (11)$$

В случае, когда значения переменных состояния отсутствуют в обучающей выборке, т.е. $\mathbf{D} = \mathbf{Y}$, оценку параметров $\hat{\boldsymbol{\theta}}$ для заданной структуры s модели позволяют найти процедуры, основанные на ЕМ-алгоритме. Этот алгоритм сходится к локальному максимуму функции правдоподобия $P(\mathbf{Y}|\boldsymbol{\theta}, s)$ и состоит из последовательности итераций, включающих Е и М-шаги. Итерационная процедура оценки параметров, согласно [3], имеет вид

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \sum_{\mathbf{X}} P(\mathbf{X}|\mathbf{Y}, \tilde{\boldsymbol{\theta}}, s) L(\mathbf{X}, \mathbf{Y}|\boldsymbol{\theta}, s), \quad (12)$$

где $\tilde{\boldsymbol{\theta}}$ и $\hat{\boldsymbol{\theta}}$ — оценки параметров на предыдущей и данной итерациях соответственно.

Соответствующую процедуру для обучения СММ представляет известный алгоритм Баума-Уэлша. На Е-шаге с использованием найденных на предшествующей итерации оценок параметров $\tilde{\boldsymbol{\theta}}$ с помощью процедуры интерполяции на фиксированном интервале, включающей рекуррентные вычисления в прямом и обратном времени, определяются так называемые прямые и обратные переменные [4]

$$\alpha(n) = P(x(n)|\mathbf{Y}_{1:n}),\ \beta(n) = P(\mathbf{Y}_{n+1:N}|x(n)),\ n = \overline{1, N}. \quad (13)$$

С помощью этих переменных вычисляются ожидаемые статистики

$$\xi_{i,j}(n) = P(x(n) = a_i, x(n+1) = a_j | \mathbf{Y}, \boldsymbol{\theta}), \quad (14)$$

$$\chi_i(n) = P(x(n) = a_i | \mathbf{Y}, \boldsymbol{\theta}). \quad (15)$$

Затем на М-шаге с использованием вычисленных величин находятся новые оценки $\hat{\boldsymbol{\theta}}$ параметров СММ. Шаги ЕМ-алгоритма повторяются до наступления сходимости итерационного процесса.

При наличии скрытых переменных структура и параметры моделей могут быть найдены с помощью структурного ЕМ-алгоритма [5], на каждой итерации которого находятся новые оценки как параметров $\hat{\boldsymbol{\theta}}$, так и структуры модели €. В рамках байесовской методологии в [6] получен вариационный байесовский ЕМ-алгоритм (VBEM), который может рассматриваться как обобщение ЕМ-алгоритма. С помощью VBEM может быть найдена оценка параметров $\hat{\boldsymbol{\theta}}$ по критерию максимума апостериорной плотности вероятности

$p(\boldsymbol{\theta} \mid \mathbf{Y}, s)$, а также нижняя граница для логарифма маргинальной функции правдоподобия $P(\mathbf{Y} \mid s)$, что позволяет выполнять селекцию моделей.

После завершения процесса обучения СММ может применяться для решения задач фильтрации и прогнозирования.

Моделирование процедур структурно-параметрического обучения

В качестве примера для моделирования рассмотрим систему с четырьмя состояниями $a_i, i = \overline{1,4}$, между которыми возможны только последовательные переходы. Так, переход от a_2 к a_4, минуя a_3, невозможен, а также запрещены обратные переходы. Такая модель (рис. 2) является частным случаем процесса гибели и размножения.

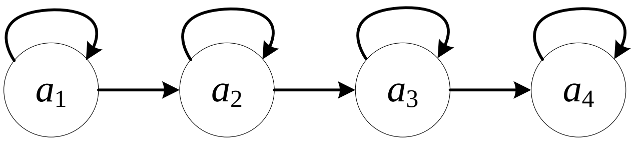


Рис. 2. Диаграмма состояний

Состояние моделируемой системы непосредственно не наблюдается. В каждый момент времени наблюдается вектор $\mathbf{y}(n) = [y_1(n), y_2(n)]^{\mathrm{T}}$, компоненты которого представляют собой реализации гауссовских случайных величин со средними μ_{1i}, μ_{2i} и дисперсиями D_{1i}, D_{2i}, зависящими от скрытого состояния a_i. Кроме того, полагается, что $y_1(n)$ и $y_2(n)$ условно независимы.

Для реализации процедур прогнозирования состояния дискретных стохастических систем в среде MatLab с использованием пакета расширения BNT разработана программа, которая позволяет проводить структурное и параметрическое обучение, а также выполнять процедуры фильтрации и прогнозирования на основе реальных наблюдений или генерируемых (синтетических) данных.

Для сравнения байесовского информационного критерия (BIC) и информационного критерия Акаике (AIC) с помощью моделирования получена зависимость оценок порядка цепи Маркова, соответствующей модели состояния СММ, от объема обучающей выборки (рис. 3). При моделировании сформированы последовательности длиной $N = 200$ элементов, сгенерированные моделью второго порядка. В каждый момент дискретного времени $n = \overline{1, N}$ с помощью AIC и BIC найдены оценки порядка марковской цепи.

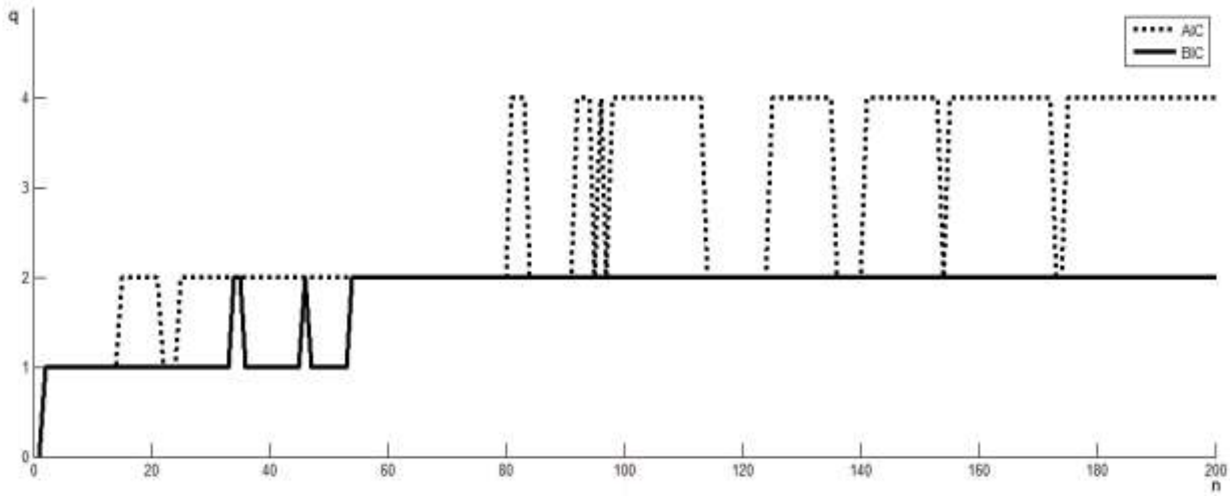


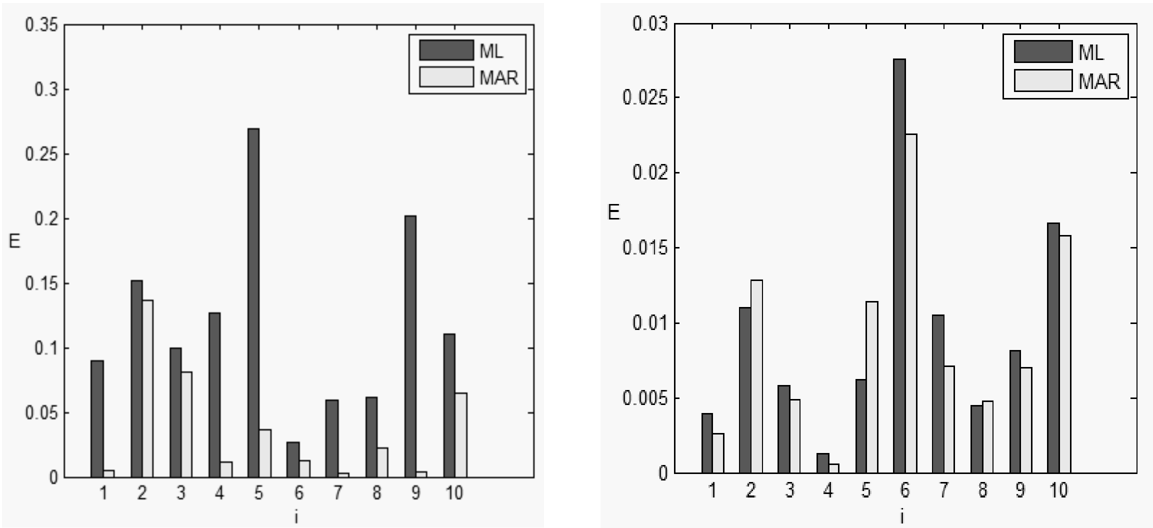
Рис. 3. Оценки порядка цепи Маркова

Результаты моделирования (рис. 3) свидетельствуют, что при малом объеме выборки решения принимаются в пользу моделей, более простых, чем модель, использованная для генерации обрабатываемых последовательностей. При увеличении длины наблюдаемой последовательности оценка порядка по BIC сходится к истинному значению. При использовании AIC наблюдается систематическое завышение порядка модели, что согласуется с выводами, представленными в [7].

При параметрическом обучении выполнено сравнение оценок параметров по критериям максимума правдоподобия (ML) и минимума среднего риска (MAR). В качестве характеристики точности оценок параметров использована ошибка обучения, которая определяется как сумма квадратов разностей между истинными значениями переходных вероятностей (известными при моделировании) и значениями, полученными в результате обучения

$$E = \sum_{i=1}^{r}\sum_{j=1}^{r}\left(A_{i,j} - \hat{A}_{i,j}\right)^2. \qquad (16)$$

На рис. 4 показаны значения ошибок, полученные при параметрическом обучении десяти цепей Маркова по выборкам различного объема.



а) *б)*

Рис. 4. Ошибки оценок переходных вероятностей по выборкам:
а – объемом $N = 10$ элементов; *б* – объемом $N = 50$ элементов

Результаты моделирования (рис. 4) демонстрируют превосходство байесовского подхода, которое в наибольшей степени проявляется в условиях малых выборок.

Моделирование процедур прогнозирования состояния системы

Для СММ (рис. 2) выполнено моделирование процедур фильтрации и прогнозирования состояния. Для этого сформирована последовательность состояний и последовательности данных наблюдения, примеры которых приведены на рис. 5. Вертикальными штриховыми линиями обозначены моменты смены состояния. Последовательность переходов a_1, a_2, a_3, a_4 можно рассматривать как представление процесса деградации системы.

Эти последовательности подаются на вход процедур фильтрации и прогнозирования: в каждый момент времени $n = \overline{1, N}$ используется часть $\mathbf{Y}_{1:n}$ сгенерированной последовательности наблюдений. С помощью фильтрации вычисляются апостериорные вероятности состояний для шагов с 1 до n, а также выполняется прогнозирование состояния системы на $\tau_{\text{П}}$ шагов. На рис. 6, *а* представлены результаты прогнозирования на $\tau_{\text{п}} = 40$ шагов в начальный момент времени, на рис. 6, *б* и 6, *в* – результаты фильтрации и прогнозирования после 40 и

80 шагов наблюдения соответственно. Графики (рис. 6) свидетельствуют, что в рассматриваемом примере процедура фильтрации позволяет достаточно точно установить смену состояний системы.

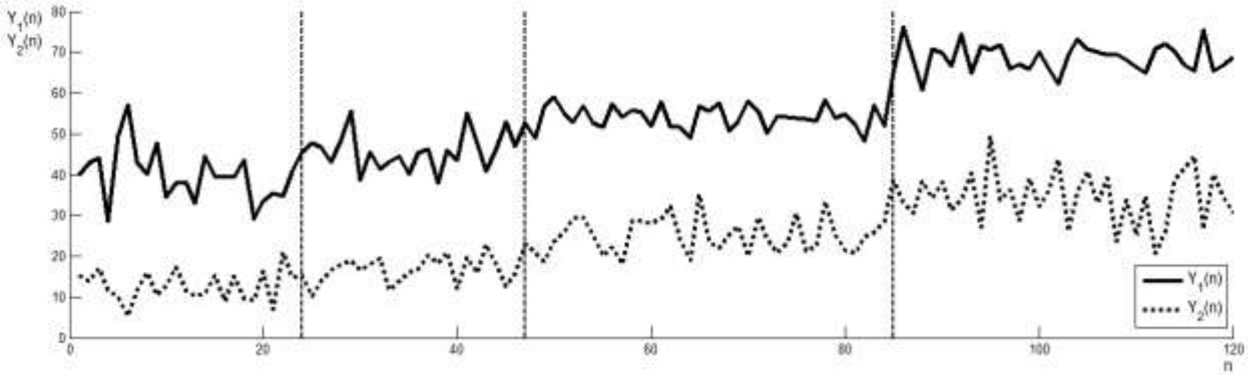
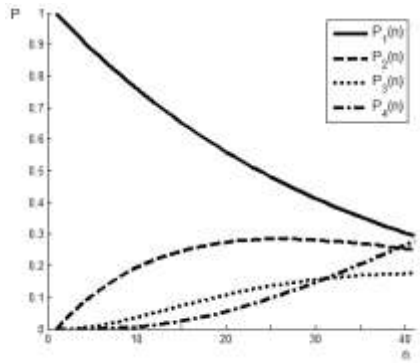
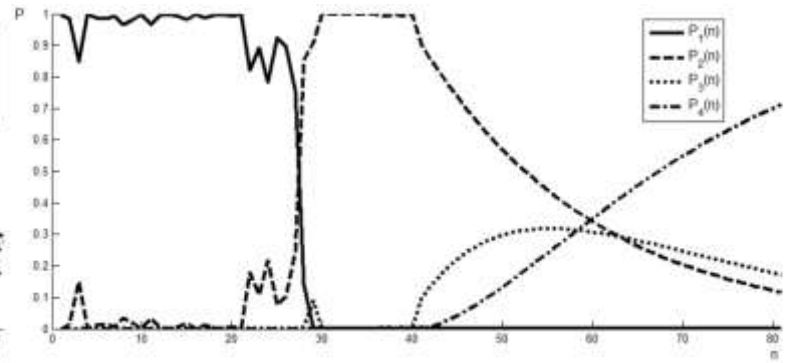


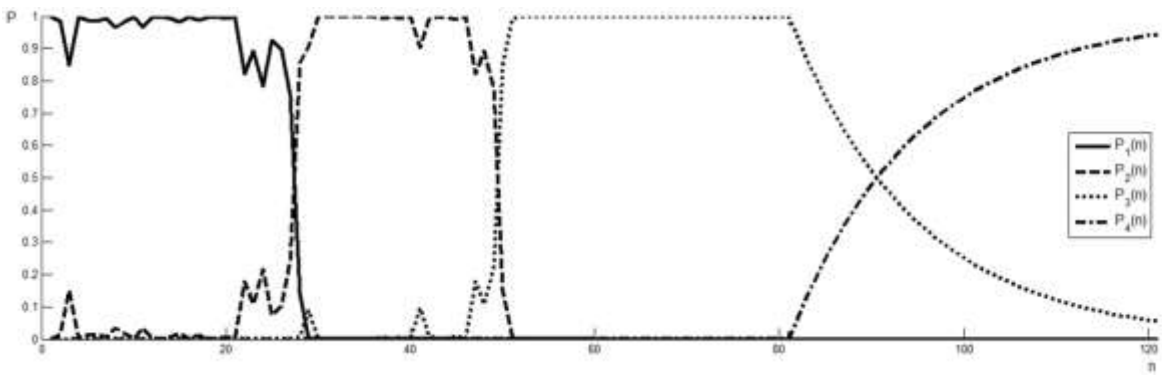
Рис. 5. Реализации данных наблюдения



а)



б)



в)

Рис. 6. Результаты прогнозирования:
а – на основе априорной информации; *б* – после 40 шагов; *в* – после 80 шагов

Вывод

На основе байесовской методологии с использованием известных алгоритмов сформированы процедуры структурно-параметрического синтеза, а также процедуры фильтрации и прогнозирования состояния стохастических систем, описываемых при помощи скрытых марковских моделей. Эти процедуры реализованы в среде MatLab в программном обеспечении прогнозирования состояния технических систем. Результаты работы могут найти применение в составе систем диагностики и управления состоянием сложных технических систем, а также систем технического обслуживания и ремонта.

Библиографический список

1. **Милов, В. Р.** Структурно-параметрический синтез нейросетевых моделей нестационарных систем / В. Р. Милов, С. А. Шалюгин // IX Всероссийская науч.-техн. конф. «Нейроинформатика-2007»: сб. науч. тр. – М.: МИФИ, 2007. Ч. 1. С. 205–213.
2. **Katz, R.** On some criteria for estimating the order of a Markov chain / Richard W. Katz // Technometrics. 1981. V. 23. № 3. – P. 243–249.
3. **Рассел, С.** Искусственный интеллект: современный подход / С. Рассел, П. Норвиг; пер. с англ. К. А. Птицына. – М.: Издательский дом «Вильямс», 2006. – 1408 с.
4. **Rabiner, L.** A tutorial on Hidden Markov Models and selected applications in speech recognition / Lawrence R. Rabiner // Proceedings of the IEEE. 1989. V. 77. № 2. – P. 257–285.
5. **Friedman, N.** The Bayesian Structural EM Algorithm // Fourteenth Conf. on Uncertainty in Artificial Intelligence. 1998. – P. 129–138.
6. **Ghahramani, Z.** Propagation algorithms for variational Bayesian learning / Z. Ghahramani, M.J. Beal // Neural Information Processing Systems 13, 2001. – P. 507–513.
7. **Csiszár, I.** The consistency of the BIC Markov order estimator / Imre Csiszár, Paul C. Shields // The Annals of Statistics. 2000. V. 28. № 6. – P. 1601–1619.

Дата поступления
в редакцию 30.03.2010

V.R. Milov, V.G. Baranov, A.Yu. Epshteyn, I.V. Shalashov

DISCRETE STOCHASTIC SYSTEM STATE PREDICTION UNDER UNCERTAINTY ON THE BASIS OF BAYESIAN APPROACH

Probabilistic-statistical approach to technical state prediction is examined. Software prototype for state prediction for the systems described by hidden Markov models is developed.

Keywords: dynamic Bayesian network, hidden Markov model, filtering, prediction, structure-parameter learning.