УДК 681.5:519.217

ЦЕЛЕСООБРАЗНОСТЬ ПРИМЕНЕНИЯ СКРЫТЫХ МАРКОВСКИХ МОДЕЛЕЙ С ЯВНЫМ ОБРАЗОМ ЗАДАННОЙ ФУНКЦИЕЙ ПЛОТНОСТИ ДЛИТЕЛЬНОСТИ СОСТОЯНИЙ В АСНИ

© 2023 ***Третьяков И.А.***

В данной работе приводится обоснование целесообразности применения специальных типов скрытых марковских моделей, в частности с явно заданной функцией плотности распределения длительности состояний, в автоматизированных системах научных исследований. Приведен алгоритм генерации последовательности наблюдений и процедура переоценки параметров для такой модели.

Ключевые слова: скрытая марковская модель, АСНИ, функция плотности, длительность состояний, переходная вероятность, вычислительная операция.

Введение. При проведении научных исследований существует постоянная потребность в анализе и обработке больших массивов данных для извлечения наиболее ценной информации. В такой роли скрытые марковские модели (СММ) стали полезным инструментом для моделирования и прогнозирования различных процессов [1-3]. СММ представляют собой статистическую модель, основанную на теории марковских процессов, и широко применяются в автоматизированных системах научных исследований (АСНИ) [4-6].

В классических типах СММ [7] принято считать наблюдения дискретными символами из конечного алфавита, а задание текущих значений этих символов выполняется посредством дискретных функций плотности вероятности. Проблема такого подхода заключается в том, что во многих прикладных задачах наблюдения являются непрерывными сигналами или векторами (акустические и речевые сигналы, радиоволны и т.д.), квантование которых зачастую приводит к большим искажениям исходного сигнала. Поэтому актуальной задачей является поиск возможностей применения специальных типов скрытых марковских моделей в автоматизированных системах научных исследований для решения прикладных задач.

В работе [8] рассмотрена применимость СММ с функцией плотности распределения наблюдений в АСНИ, в которых наблюдения O соответствуют скрытым состояниям S модели. Для такой СММ вероятность того, что она останется в текущем состоянии d моментов времени (длительность состояний), определяется как вероятность последовательности наблюдений $O = \left\{ \underset{1}{S_i}, \underset{2}{S_i}, \underset{3}{S_i}, \ldots, \underset{d}{S_i}, \underset{d+1}{S_j} \neq S_i \right\}$ модели λ, отсюда:

$$P(O \mid \lambda) = (a_{ij})^{d-1}(1 - a_{ij}) = p_i(d), \qquad (1)$$

где a_{ij} - переходная вероятность (вероятность смены состояний).

Величина $p_i(d)$ в формуле (1) представляет собой дискретную функцию плотности вероятности пребывания модели λ в состоянии i в течение времени d. Эта функция является экспоненциальной и характеризует длительность данного состояния в цепи Маркова [9].

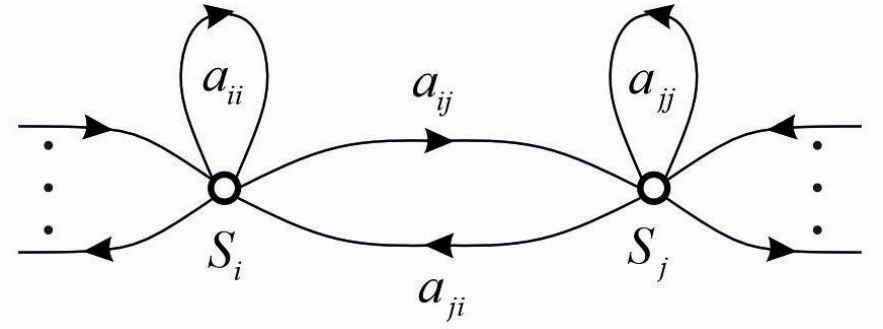
Математическое ожидание непрерывного пребывания в текущем состоянии (ожидаемое число повторений текущего состояния) из (1) определяется следующим образом:

$$\bar{d}_i = \sum_{d=1}^{\infty} d p_i(d) = \sum_{d=1}^{\infty} d (a_{ij})^{d-1}(1-a_{ij}) = \tfrac{1}{1-a_{ij}} \cdot$$

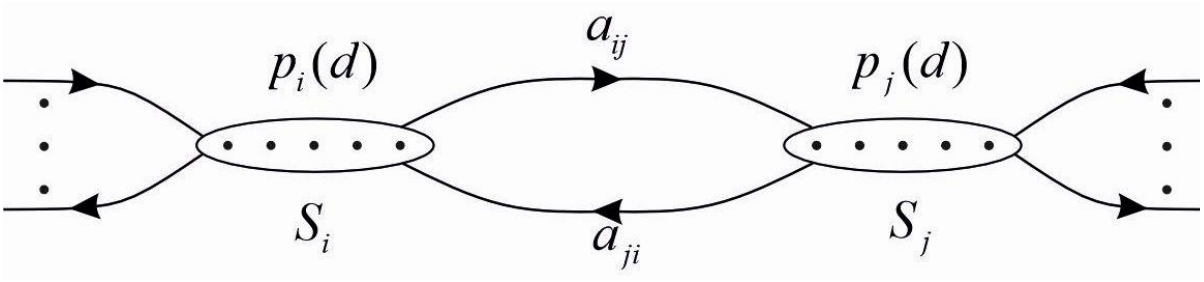
Однако, такой способ моделирования длительности состояний не является универсальным, например при распознавании образов и в некоторых других прикладных случаях [4, 5] выходит целесообразнее моделировать ее явно. Таким образом, в работе рассматривается целесообразность применения специальных типов скрытых марковских моделей, в частности с явным образом заданной функцией плотности длительности состояний, в АСНИ.

Модели с функцией плотности длительности состояний. Впервые СММ с функциями плотности вероятности длительности состояний введены и рассмотрены в работах [10, 11] по базовым принципам распознавания речи.

На рисунке 1 показаны очевидные отличия между скрытыми марковскими моделями с неявно и явно заданными функциями плотности длительности состояния для двух скрытых состояний модели S_i и S_j.



а)



б)

Рис. 1. Визуализация отличий СММ:

а) классическая СММ

б) СММ с явно заданной плотностью длительности состояний

Для модели (рисунок 1а) функции плотности длительности скрытых состояний S определяются по переходным вероятностям a и являются экспоненциальными (1), т.е. заданы не явно.

Для модели (рисунок 1б) вероятности перехода в то же состояние a_{ii} и a_{jj} приняты равными нулю, и соответственно функции плотности длительности состояний $p_i(d)$ заданы явно. Иными словами, в такой модели не допускаются переходы в то же текущее состояние, а переход из текущего состояния в другое будет происходить только после генерации такого количества наблюдений O, которое задано функцией плотности $p_i(d)$. Таким образом, в СММ с явно заданной $p_i(d)$ генерация последовательности наблюдений O осуществляется посредством следующего алгоритма:

- состояние в первый момент времени $q_1 = S_i$ определяется в соответствии с матрицей вероятности начальных состояний π_i;

- длительность нахождения в первом состоянии d_1 определяется в соответствии с функцией плотности $p_{q_1}(d)$;

- вводится ограничение D максимального значения d_1;

- наблюдения $\left(O_1, O_2, \ldots, O_{d_1}\right)$ выбираются из матрицы совместных наблюдений b_{q_1};

- проверяется условие независимости выбранных наблюдений $b_{q_1}(O_1, O_2, ..., O_{d_1}) = \prod_{t=1}^{d_1} b_{q_1}(O_t)$, где O_t - наблюдение в момент времени t;

- состояние в следующий момент времени $q_2 = S_j$ определяется в соответствии с переходной вероятностью $a_{q_1 q_2}$;

- переходная вероятность устанавливается $a_{q_1 q_2} = 0$ для недопущения возврата модели в то же состояние.

Далее, для применения к такой модели методов повторной оценки параметров и вычисления $P(O|\lambda)$ необходимо, чтобы интервалы длительности состояний были полностью включены в последовательность наблюдений, т.е. необходимо ввести следующие уточнения:

- первое состояние заканчивается в момент времени $t = 1$;

- последнее состояние заканчивается в момент времени $t = T$;

- за время t первых наблюдений сменилось r состояний $\left(q_1, q_2, \ldots, q_r\right)$ с длительностями $\left(d_1, d_2, \ldots, d_r\right)$, т. е. $q_r = S_i$, $\sum_{s=1}^{t} d_s = t$.

Таким образом, прямые и обратные переменные для применения алгоритма прямого-обратного хода [12] определяются следующим образом:

$$\alpha_t(i) = P\left(O_1, O_2, \ldots, O_t, S_i \text{ заканчивается в момент } t \middle| \lambda\right),$$

$$\alpha_t^*(i) = P(O_1, O_2, ..., O_t, S_i \text{ начинается в момент } t+1 | \lambda),$$

$$\beta_t(i) = P(O_{t+1}, ..., O_T | S_i \text{ заканчивается в момент } t, \lambda),$$

$$\beta_t^*(i) = P(O_{t+1}, ..., O_T | S_i \text{ начинается в момент } t+1, \lambda).$$

Таким образом, формулы повторной оценки параметров для модели с явно заданной плотностью длительности состояний имеют вид:

$$\overline{\pi_i} = \frac{\pi_i \beta_0^*(i)}{P(O \mid \lambda)}; \ \overline{a_{ij}} = \frac{\sum_{t=1}^{T} \alpha_t(i) a_{ij} \beta_t^*(j)}{\sum_{j=1}^{N} \sum_{t=1}^{T} \alpha_t(i) a_{ij} \beta_t^*(j)};$$

$$\overline{b_i}(k) = \frac{\sum_{t=1, O_t=k}^{T} \left[\sum_{T<t} \alpha_T^*(i) \beta_T^*(i) - \sum_{T<t} \alpha_T(i) \beta_T(i) \right]}{\sum_{k=1}^{M} \sum_{t=1, O_t=V_k}^{T} \left[\sum_{T<t} \alpha_T^*(i) \beta_T^*(i) - \sum_{T<t} \alpha_T(i) \beta_T(i) \right]};$$

$$\overline{p_i}(d) = \frac{\sum_{t=1}^{T} \alpha_t^*(i) p_i(d) \beta_{t+d}(i) \prod_{s=t+1}^{t+d} b_i(O_s)}{\sum_{d=1}^{D} \sum_{t=1}^{T} \alpha_t^*(i) p_i(d) \beta_{t+d}(i) \prod_{s=t+1}^{t+d} b_i(O_s)}.$$

Физический смысл формул повторной оценки заключается в следующем:

- $\overline{\pi_i}$ - вероятность того, что при заданных наблюдениях O первым состоянием модели было состояние S_i;
- $\overline{a_{ij}}$ - ожидаемое число переходов из состояния S_i в состояние S_j при условии, что прямые переменные, для которых состояние S_i заканчивается в момент t, объединяются с обратными переменными, для которых новое состояние Sj начинается в момент $t+1$;
- $\overline{b_i}(k)$ - ожидаемое число появлений символа наблюдения $O_t = V_k$ в состоянии S_i, нормированное ожидаемым числом появления любых символов в состоянии S_i;
- $\overline{p_i}(d)$ - отношение ожидаемого числа раз, в которых состояние S_j имело длительность d, к общему числу появлений состояния S_j любой длительности.

Целесообразность применения в АСНИ. Целесообразность явного задания функций плотности вероятности длительности состояний и применения таких СММ в автоматизированных системах научных исследований обосновывается значительным повышением качества и точности моделирования, например до 92% при распознавании речи [13], при распознавании образов и в некоторых других прикладных случаях [4, 5]. Однако, повышение эффективности модели приводит к увеличению вычислительных мощностей. Например, из-за смены длительностей состояний для определения значений переменных алгоритма прямого-обратного хода, объем требуемой для вычислений памяти увеличивается в $\sim D$ раз, а количество вычислительных операций в $\sim \frac{D^2}{2}$ раз в сравнении с классическими СММ. Таким образом, для усредненного значения при выполнении задач, связанных с обработкой речи $D = 50$ количество вычислительных операций увеличивается более чем в 10^3 раз.

Этот недостаток можно компенсировать посредством параметрического задания функции плотности длительности состояний, например семейство гауссовских плотностей распределения $p_i(d) = \chi(d, \mu_i, \sigma_i^2)$ с параметрами μ_i, σ_i^2 или набор плотностей, описываемых гамма-функцией $p_i(d) = \frac{\eta_i^{v_i} d^{v_i - 1} e^{-\eta_i d}}{\Gamma(v_i)}$, с параметрами v_i, η_i, средним значением $v_i \eta_i^{-1}$ и дисперсией $v_i \eta_i^{-2}$. Также к сокращению вычислительных операций приводит равномерное распределение длительности состояний и применение алгоритма декодирования Витерби для ограниченного пути [14].

Еще одним недостатком такой СММ является необходимость помимо обычных параметров повторно оценивать дополнительное, равное D количество новых параметров. Это компенсируется посредством ограничения количества наблюдений T, и тогда обучающее множество модели содержит даже меньшее количество смен состояний и меньшее количество данных для оценивания $p_i(d)$, чем классическая модель.

Кроме того, для автоматизированных систем научных исследований, в которых в качестве СММ применяется лево-правая модель (модель Бакиса) [7] с количеством состояний пропорциональным средней длительности последовательности наблюдений, задание функции плотности длительности состояния явным образом не является ни необходимым, ни тем более целесообразным.

Выводы. Использование СММ с явно заданной функцией плотности длительности состояний в АСНИ позволяет получить наиболее важную информацию из сложных массивов данных, повысить точность и эффективность моделирования, улучшить качество распознавания в сравнении с классическими моделями.

Таким образом, специальные типы скрытых марковских моделей функцией плотности длительности состояний вполне применимы в автоматизированных системах научных исследований для решения широкого спектра задач.

СПИСОК ИСТОЧНИКОВ

1. Заморенов, М. В. Моделирование полумарковской системы с учетом непростого потока восстановления / М. В. Заморенов, В. Я. Копп, Н. И. Чаленков // Известия Тульского государственного университета. Технические науки. – 2023. – № 3. – С. 112-119. – DOI 10.24412/2071-6168-2023-3-112-120. – EDN FVTFDF.
2. Обжерин, Ю. Е. Полумарковская и скрытая марковская модели функционирования автономного ветродизельного комплекса / Ю. Е. Обжерин, С. М. Сидоров, С. Н. Федоренко // Информационные и математические технологии в науке и управлении. – 2022. – № 4(28). – С. 70-82. – DOI 10.38028/ESI.2022.28.4.005. – EDN BSUKZZ.
3. Токарев, В. Л. Скрытые марковские модели в задаче обнаружения атак на компьютерные сети / В. Л. Токарев // Чебышевский сборник. – 2021. – Т. 22, № 5(81). – С. 391-399. – DOI 10.22405/2226-8383-2021-22-5-391-399. – EDN FDRRVI.
4. Разработка автономной автоматизированной системы непрерывного контроля состояния строительного объекта / В. Ю. Подлесный, В. И. Тимченко, И. А. Третьяков, В. В. Данилов // Вестник Донецкого национального университета. Серия Г: Технические науки. – 2023. – № 2. – С. 51-59. – EDN PHCERH.
5. Тимченко, В. И. Определение параметров упругой среды при ударном воздействии / В. И. Тимченко, В. А. Сорока, В. А. Ступак [и др.] // Вестник Донецкого национального университета. Серия Г: Технические науки. – 2020. – № 4. – С. 18-25. – EDN UNGIDR
6. Третьяков, И. А. Исследование применимости автоматизированной системы лингвистического анализа экспериментальных данных научных исследований / И. А. Третьяков // Интеллектуальные информационные системы: труды Международной научно- практической конференции (Воронеж, 2-4

декабря 2020 г.): в 2 ч. / ФГБОУ ВО «Воронежский государственный технический университет». − Воронеж: Изд-во ВГТУ, 2021. − Ч.2. – С. 38-41. – EDN WLOAWZ.
7. Третьяков, И. А. Классификация скрытых марковских моделей / И. А. Третьяков // Донецкие чтения 2022: образование, наука, инновации, культура и вызовы современности: Материалы VII Международной научной конференции, посвящённой 85-летию Донецкого национального университета (Донецк, 27–28 октября 2022 г.). – Том 2: Физические, технические и компьютерные науки / под общей редакцией проф. С.В. Беспаловой. – Донецк: Изд-во ДонНУ, 2022. – С. 203-205. – EDN RSVCWV.
8. Третьяков, И. А. Обоснование применения скрытых марковских моделей с функцией плотности распределения наблюдений в АСНИ / И. А. Третьяков // Вестник Донецкого национального университета. Серия Г: Технические науки. – 2023. – № 2. – С. 16-21. – EDN DOIFMY.
9. Baum, L. E. Statistical inference for probabilistic functions of finite start Markov chains / L. E. Baum, T. Petrie // The annals of mathematical statistics. – 1966. – V. 3. – № 6. – P. 1554-1563. – DOI 10.1214/aoms/1177699147.
10. Levinson, S. E. An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition / S. E. Levinson, L. R. Rabiner, M. M. Sondhi // The Bell System Technical Journal. – 1983. – V. 62. – № 4. – P. 1035-1074. – DOI 10.1002/j.1538-7305.1983.tb03114.x.
11. Juang, B. H. On the hidden Markov model and dynamic time warping for speech recognition — A unified view / B. H. Juang // AT&T Bell Laboratories Technical Journal. – 1984. – V. 63. – № 7. – P. 1213-1243. – DOI 10.1002/j.1538-7305.1984.tb00034.x.
12. Yu, S. Z. An Efficient Forward–Backward Algorithm for an Explicit - Duration Hidden Markov Model / S. Z. Yu, H. Kobayashi // IEEE Signal Processing Letters. – 2003. – Vol. 10. – № 1. – P. 11-14. – DOI 10.1109/LSP.2002.806705.
13. Третьяков, И. А. Реализация и тестирование автоматизированной системы текстонезависимой идентификации диктора / И. А. Третьяков, Е. Н. Кожекина // Импортозамещение, научно-техническая и экономическая безопасность: сб. ст. V Междунар. науч.-техн. конф. «Минские научные чтения-2022» в 3 т. (Минск, 07–09 декабря 2022 г.). – Т. 2. – Минск: БГТУ, 2022. – С. 338-342. – EDN ENMOZX.
14. Forney, D. G. The Viterbi Algorithm / D. G. Forney // Proceedings of the IEEE. – 1973. – Vol. 61. – № 3. – P. 268-278. – DOI 10.1109/PROC.1973.9030.

Поступила в редакцию 11.08.2023 г., рекомендована к печати 28.08.2023 г.

THE EXPEDIENCY OF USING HIDDEN MARKOV MODELS WITH AN EXPLICITLY GIVEN FUNCTION OF THE DENSITY OF THE DURATION OF STATES IN ASRS

Tretiakov I.A.

This paper provides a justification for the expediency of using special types of hidden Markov models, in particular with an explicitly given density function of the distribution of the duration of states, in automated systems of scientific research. An algorithm for generating a sequence of observations and a procedure for revaluation of parameters for such a model is given.

***Keywords*:** hidden Markov model, ASRS, density function, duration of states, transient probability, computational operation.

Третьяков Игорь Александрович
кандидат технических наук, заместитель декана по научной работе, доцент кафедры радиофизики и инфокоммуникационных технологий ФГБОУ ВО «Донецкий государственный университет»,
Российская Федерация, ДНР, г. Донецк.
E-mail: i.tretiakov@mail.ru

Tretiakov Igor Aleksandrovich
Candidate of Technical Sciences, Deputy Dean for Research, Associate Professor at Department of Radiophysics and Infocommunication Technologies of Donetsk State University,
Russian Federation, DPR, Donetsk.