

УДК 519.6

ПРИМЕНЕНИЕ МАРКОВСКИХ МОДЕЛЕЙ К ПРОБЛЕМЕ РАСПОЗНАВАНИЯ ЗАСЫПАНИЙ ЧЕЛОВЕКА

Д. В. Пархоменко
(МГУ им. М. В. Ломоносова)

Рассматривается применение методики вероятностных автоматов к проблеме распознавания засыпаний человека за рулем. Описана вероятностная модель моргания человека и изучены ее некоторые простейшие свойства. Представлен алгоритм определения засыпания и освещены результаты тестирования построенной аппаратно-программной реализации алгоритма.

Введение

Разработки систем, не позволяющих водителю транспортного средства засыпать, ведутся крупнейшими автопроизводителями уже более 40 лет. Но только сейчас качество таких систем достигло приемлемого уровня надежности и удобства использования. Требования к системе, отвечающей за безопасное вождение, очень высоки: она должна быть устойчива к постоянным и непостоянным возмущителям, не должна допускать ошибок, должна быть простой в эксплуатации и не причинять дискомфорта водителю.

Наиболее удовлетворяющей всем этим требованиям оказалась система, основанная на распознавании глаза водителя с помощью видеокамер. Большинство систем, выпускаемых в настоящее время, основано именно на этом принципе.

Известно о наличии связей степени бодрствования человека с процессом его моргания (интенсивность морганий, длительность морганий, ширина открытости глаза и др.). Поэтому восстребованной является задача построения математической модели процесса моргания и последующего его изучения.

В данной работе рассматривается применение алгоритма, широко используемого для распознавания речи [1], к проблеме моделирования и распознавания морганий человека.

Постановка задачи

Итак, имеется видеоизображение человеческого глаза — набор кадров. Пусть задана какая-

либо числовая характеристика O степени открытости глаза человека. Припишем эту числовую характеристику каждому кадру. Тем самым имеем последовательность чисел O_1, \dots, O_T , где T — число кадров. Без ограничения общности будем считать, что $O_i \in 0, \dots, 9$.

Пусть имеется несколько таких последовательностей из чисел. Тогда требуется построить такой механизм, который в процессе своего функционирования выдавал бы последовательности чисел, максимально похожие в том или ином смысле на заданные последовательности, а заданные последовательности выдавал бы с максимальной долей вероятности.

В качестве такого механизма хорошо выбрать вероятностный автомат [2] — набор (Q, E, A, B, π) , где Q — множество состояний q_1, \dots, q_N ; E — множество выходных символов, которое здесь будет представлять собой набор целых чисел от 0 до 9; $A = \{a_{ij}\}$ — матрица переходов, $1 \leq i, j \leq N$; $B = \{b_{jk}\}$ — матрица выходов, $1 \leq k \leq 10$; $\pi = \{\pi_i\}$ — распределение вероятностей выбора начального состояния.

Покажем, как вероятностный автомат может порождать последовательности чисел O_1, \dots, O_T .

На первом шаге выбираются начальное состояние в соответствии с распределением вероятностей выбора начального состояния π^1 и выходной символ O_1 .

¹Всюду ниже будем полагать, что $\pi = \{1, 0, \dots, 0\}$. Иными словами, автомат начинает работу всегда в первом состоянии.

Выходной символ выбирается с помощью матрицы $B = \{b_{jk}\}$, где $b_{jk} = b_j(k) = P(O_t = k | t \rightarrow q_j)$ — вероятность выдать число k , находясь в состоянии j . Далее используется матрица A , где $a_{ij} = P(t \rightarrow q_j | t-1 \rightarrow q_i)$ — условная вероятность того, что в следующий момент времени автомат будет находиться в j -м состоянии при условии, что в данный момент времени он находится в состоянии i . В соответствии с матрицей A осуществляется переход в следующее состояние, с помощью B находится следующий выходной символ (число от 0 до 9) и т. д. Выполнив T шагов, получим последовательность O_1, \dots, O_T .

Алгоритм решения

Процесс моргания человека предлагается рассмотреть как результат функционирования двух вероятностных автоматов $\mathcal{J} = (E, \Phi)$, $\mathcal{R} = (K, \Lambda)$, где матрицы E , K — матрицы переходов; Φ , Λ — матрицы выходов. Автомат \mathcal{J} моделирует степень открытости глаза, когда моргание отсутствует, и определяет, когда будет моргание; \mathcal{R} моделирует это моргание. Схематично результат моделирования можно представить в виде рис. 1.

Для решения поставленной задачи будем применять алгоритм Баума—Уэлша [3] обучения вероятностных автоматов. Алгоритм хорошо зарекомендовал себя в задачах распознавания речи,

имеет высокую скорость сходимости и приемлем для вычисления.

Алгоритм основан на определении функции правдоподобия $P(O_1, \dots, O_T | A, B, \pi)$ — вероятности выдать заданную последовательность при фиксированных матрицах перехода, выхода и векторе начального распределения вероятностей. Эта функция в общем случае может быть вычислена за N^T операций, где N — число состояний рассматриваемого вероятностного автомата, T — длина наблюдаемой последовательности.

Для построения более эффективной вычислительной процедуры Баум ввел прямую и обратную функции частичного правдоподобия

$$\alpha_t(i) = P(O_1, \dots, O_t, t \rightarrow q_i | A, B, \pi);$$

$$\beta_t(i) = P(O_{t+1}, \dots, O_T | t \rightarrow q_i, A, B, \pi).$$

В начальный момент положим

$$\alpha_1(1) = b_1(O_1), \quad \alpha_1(i) = 0 \quad \forall i \neq 1;$$

$$\beta_T(i) = 1, \quad 1 \leq i \leq N.$$

Рекурсивно будем вычислять остальные правдоподобия, исходя из очевидных формул:

$$\alpha_{t+1}(j) = \left(\sum_{i=1}^N \alpha_t(i) a_{ij} \right) b_j(O_{t+1}),$$

$$1 \leq t \leq T-1, \quad 1 \leq j \leq N;$$

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \quad 1 \leq i \leq N.$$

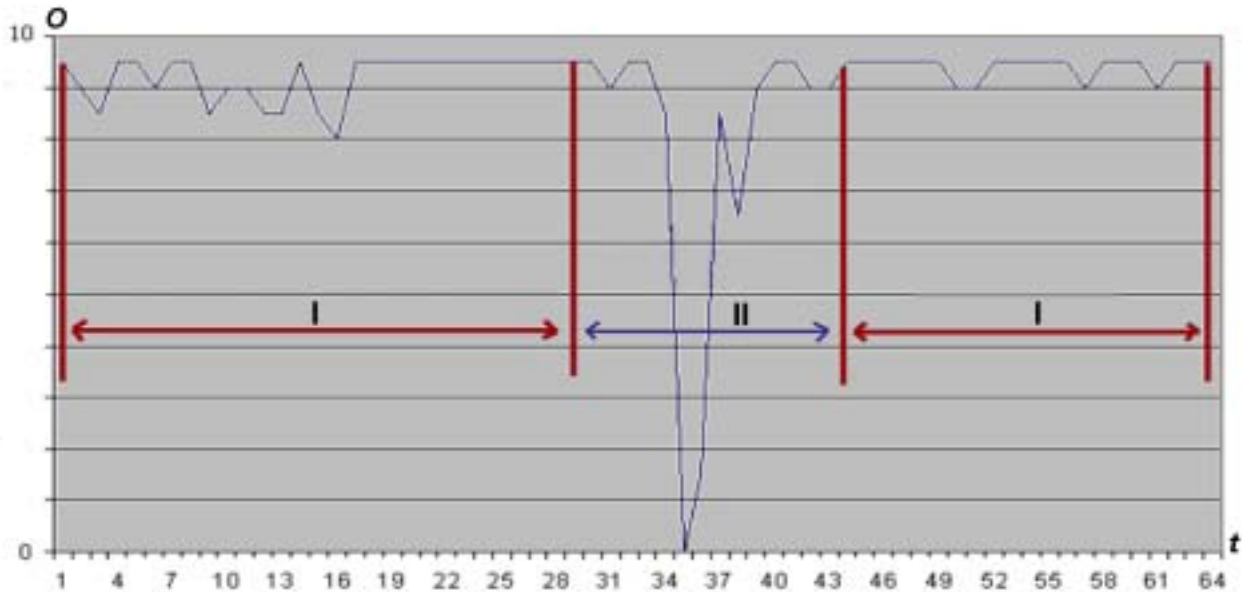


Рис. 1. Изображение части одной из последовательностей O_1, \dots, O_T : I — моделируется первым автоматом; II — моделируется вторым автоматом

Тогда рассматриваемая функция правдоподобия может быть вычислена по формуле

$$P(O_1, \dots, O_T | A, B, \pi) = \sum_{i=1}^N \alpha_T(i).$$

Тем самым правдоподобие вычисляется за TN^2 операций.

Для вычисления новых матриц переходов и выходов потребуется введение еще одной переменной

$$\xi_t(i, j) = P(t \rightarrow q_i, t+1 \rightarrow q_j | O, A, B, \pi),$$

которая является вероятностью того, что при заданной последовательности наблюдений $O = O_1, \dots, O_T$ автомат в моменты времени t и $t+1$ будет находиться соответственно в состояниях q_i и q_j . Используя прямую и обратную функции правдоподобия, определенные выше, можно записать

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{P(O | A, B, \pi)}.$$

Положим $\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j)$, где $1 \leq i, j \leq N$, $t = 1, \dots, T$. $\gamma_t(i)$ — апостериорная вероятность того, что при заданной последовательности наблюдений O автомат в момент времени t будет находиться в состоянии q_i . Тогда новое значение элемента \hat{a}_{ij} матрицы перехода \hat{A} равно отношению числа ожидаемых переходов из состояния q_i к числу ожидаемых переходов из состояния q_i в q_j , т. е.

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(j)}.$$

Формула переоценки элементов $\hat{b}_j(k)$ матрицы выходов \hat{B} будет выглядеть таким образом:

$$\hat{b}_j(k) = \frac{\sum_{\substack{t=1 \\ O_t=k}}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)}.$$

Баум показал [4], что новые оценки, полученные из вышеприведенных формул, лучше предшествующих в том смысле, что

$$P(O_1, \dots, O_T | A, B, \pi) < P(O_1, \dots, O_T | \hat{A}, \hat{B}, \pi).^2$$

²Вектор π не изменялся.

Кроме того, повторное применение формул переоценки матриц A и B в пределе при $T \rightarrow \infty$ обеспечивает сходимость параметров автомата к оценкам максимального правдоподобия.

Как показали эксперименты, человек не может моргнуть с такой скоростью, чтобы закрытый глаз в кадре с номером N стал открытым в кадре $N+1$.³ Между моментами закрытого и открытого глаза обязательно должен быть момент полукрытого глаза. Но закрыть глаз со скоростью, достаточной для того, чтобы камера не уловила момент полукрытого глаза, человек может.

Основываясь на этих наблюдениях, можно представить физически корректное первоначальное приближение искомого автомата $\mathcal{J} = (E, \Phi)$ в виде схемы, изображенной на рис. 2. На схеме обозначены состояния автомата и возможные переходы между ними. Неизображенным переходам соответствует нулевая вероятность перехода. Вероятности переходов $p1, q1, p2, q2, p3, q3, p4, q4$ определяются с помощью программной реализации алгоритма Баума—Уэлша.

Если автомат находится в первом состоянии, это означает, что глаз у человека открыт, во втором — полукрыт, в третьем — закрыт, в четвертом — полукрыт, в пятом — открыт. При этом в каждом из пяти состояний характеристика степени открытости глаза может принимать любое значение от 0 до 9, конечно, с учетом вероятности.

Первоначально вероятности переходов $p1, q1, p2, q2, p3, q3, p4, q4$ между состояниями и вероятности выходов корректируются вручную, исходя из сравнения видеозаписи моргания и графика чисел-характеристик. Это позволяет сразу приблизиться к глобальному максимуму функции правдоподобия и избежать громоздких вычислений в последующем его нахождении.

После нахождения начальных значений матриц выходов и переходов автомата к нему применяется алгоритм обучения Баума—Уэлша. В данной работе число итераций, равное 2, было выбрано исходя из соображений "привыкания" автомата к обучающей последовательности. Если слишком долго обучать автомат на одной последовательности, то он настраивается на нее и не дает качественного результата. При попытке обучить его на другой последовательности авто-

³Предполагается использование камер со скоростью записи не менее 24 кадров в секунду.

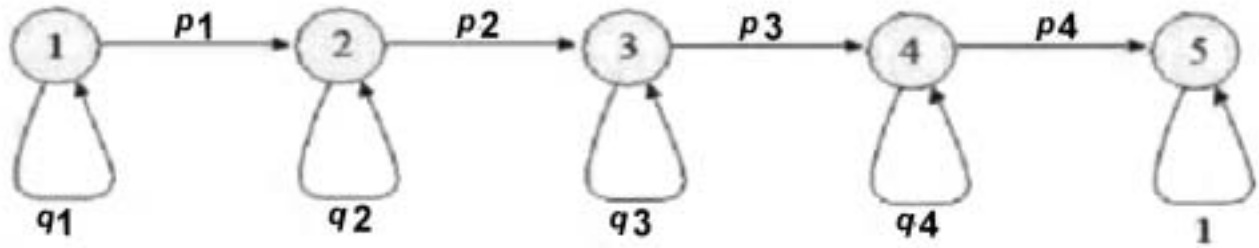


Рис. 2. Схема начального приближения автомата $\mathcal{J} = (E, \Phi)$

мат придется полностью переучивать и все предыдущие его характеристики будут уничтожены, следовательно, результат слишком качественного обучения на двух различных последовательностях превратится в результат обучения по последней последовательности. Динамика роста вероятности P выдать заданную последовательность, представленную на рис. 3, в зависимости от числа итераций алгоритма Баума—Уэлша такова:⁴

- 1 итерация — $P = 3,40285 \cdot 10^{-8}$;
- 2 итерации — $P = 3,32628 \cdot 10^{-7}$;

- 3 итерации — $P = 9,13647 \cdot 10^{-7}$;
- 4 итерации — $P = 1,0837 \cdot 10^{-6}$;
- 10 итераций — $P = 1,09698 \cdot 10^{-6}$;
- 500 итераций — $P = 1,09699 \cdot 10^{-6}$.

Отсюда видно, что брать более трех итераций нецелесообразно.

Далее автомат $\mathcal{J} = (E, \Phi)$ снова обучается, но уже на другой последовательности. Затем это повторяется необходимое число раз. В данной работе автомат обучался 6 раз на последовательностях длиной 20 (максимальная длина моргания в экспериментах не превышала 20 кадров).

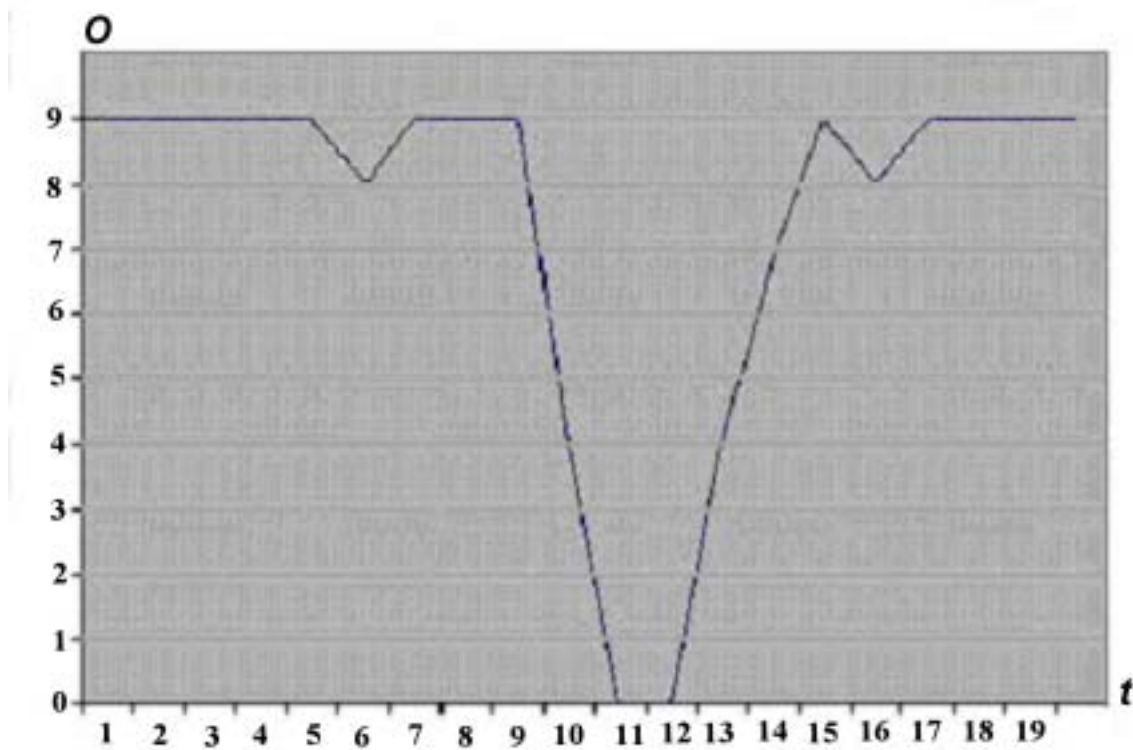


Рис. 3. График рассматриваемой последовательности

⁴На остальных последовательностях наблюдается подобная ситуация.

Построение автомата $\mathcal{R} = (K, \Lambda)$, описывающего поведение человека между морганиями, мало чем отличается от построения автомата $\mathcal{J} = (E, \Phi)$, за исключением того, что автомат \mathcal{R} выдает последовательности, состоящие из 0 и 9. Если выпадает 0, значит, глаз "достаточно закрыт" и надо подключать автомат \mathcal{J} . Автомат \mathcal{R} был обучен на 10 различных последовательностях длиной 300. Количество итераций алгоритма Баума—Уэлша полагалось равным 5.

Полученная система из двух автоматов функционирует следующим образом. Работа начинается с автомата \mathcal{R} . Выдавая 9 или 0, он определяет, будет ли в данный момент моргание или глаз будет открыт.⁵ Если моргание определено, то 10 предшествующих нулю и последующих за ним чисел заменяются на результат работы автомата \mathcal{J} . Зависимость полученной таким образом последовательности символов (точнее ее части) от времени представлена на рис. 4 (см. также цветную вкладку).

Аналогичный по своей сути подход был применен и для распознавания морганий. Автомат $\mathcal{R} = (K, \Lambda)$ теперь интерпретировался как автомат "неморгания", т. е. он реализовывал последовательности длиной 20, на которых моргания не было. Автомат был обучен в соответствии с алгоритмом Баума—Уэлша на 10 различных последовательностях.

Имея автоматы моргания и неморгания и входную последовательность, легко определить, было ли на ней моргание или нет. Входной сигнал с видеокамер преобразовывался в сигнал, состоящий из целых чисел от 0 до 9. Далее из сигнала выбирались последние 20 чисел и определялось, какой автомат выдаст данную последовательность с наибольшей долей вероятности $P(O_1, \dots, O_T | A, B, \pi)$. В случае, если вероятность выдачи заданной последовательности автоматом моргания была больше, то считалось, что моргание произошло.

Было установлено, что такой подход к определению морганий у человека работает точнее, нежели использованный ранее для данной задачи подход пороговых функций, при котором моргание считалось произошедшим, если входной сигнал в данный момент времени не превышал некоторого порога, уникального для данного периода времени. Эта точность возникает за счет того, что подход, основанный на использовании автоматов, устойчив к помехам разного рода, будь то плавное снижение среднего значения входной последовательности ниже порога или резкий скачок к 0.⁶

Автоматный подход реализован и опробован. При этом неправильное определение моргания, т. е. ошибка, имеет место только в одном случае на записи в 15 минут,

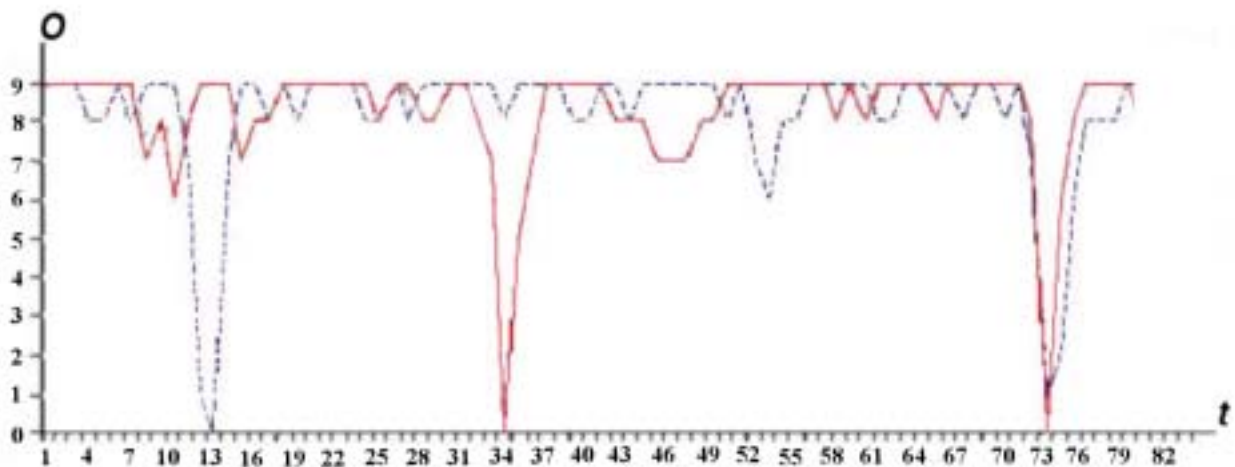


Рис. 4. Сравнительный график обучающей (---) и выходной (—) последовательностей

⁵Достижение 0 автоматом \mathcal{R} означает, что моргание произошло.

⁶Использование порогового алгоритма на последовательностях такого рода приводит к ошибке.

в то время как при пороговом подходе — в среднем в пяти случаях на такой же записи.

Вышеописанный алгоритм позволяет с высокой точностью смоделировать поведение числовых характеристик степени открытости глаза и, как следствие, может заменить проведение повторных экспериментов, а также позволит определять моргания с более высокой точностью в реальном времени.

Автор выражает благодарность доктору физ.-мат. наук, профессору Д. Н. Бабиню, а также сотрудникам механико-математического факультета МГУ кандидату физ.-мат. наук И. Л. Мазуренко и А. Б. Холоденко за помощь и ценные советы в работе, часть которой освещена в статье.

Список литературы

1. *Huang X. D., Ariki Y., Jack M. A.* Hidden Markov Models for Speech Recognition. Edinburgh University Press, 1990.
2. *Бухараев Р. Г.* Основы теории вероятностных автоматов. М.: Наука, 1985.
3. Левинсон С. Э. Структурные методы автоматического распознавания речи // ТИИЭР. 1985. Т. 73, № 11.
4. *Rabiner L., Juang B. H.* Fundamentals of Speech Recognition. Prentice Hall, 1995.

Статья поступила в редакцию 11.03.08.
