

Моделирование вероятностными источниками

Д. В. Пархоменко

Скрытая марковская модель (СММ) — это марковская цепь, обладающая видимым пользователю «выходом». В каждом состоянии, в соответствие с зависящим от состояния законом распределения, СММ выдает выходной символ и переходит в другое состояние как марковская цепь. В случае конечного выходного алфавита СММ называют вероятностным источником. В статье описано применение аппарата скрытых марковских моделей и рассказано об актуальных инженерных задачах, где они используются.

Ключевые слова: скрытые марковские модели, распознавание образов, вероятностный источник.

1. Вероятностный источник

Под вероятностным источником (вероятностным автоматом без входа) понимается набор $L = (Q, E, A, B, \pi)$, где Q — конечное множество состояний q_1, \dots, q_N , E — конечное множество выходных символов, матрица $A = \{a_{ij}\}$, a_{ij} — это вероятность перехода из состояния q_i в состояние q_j , $1 \leq i, j \leq N$, $B = \{b_{ik}\}$, b_{ik} — это вероятность выдать символ ε_k в состоянии q_i , где $1 \leq k \leq |E|$, $\pi = \{\pi_i\}$ — вектор распределения вероятностей начального состояния, $1 \leq i \leq N$ [1, 2].

Вероятностный источник работает по тактам и порождает слова из E^* . Сначала, в соответствие с распределением π , выбирается начальное состояние q_i , и, в соответствии с матрицей $\{b_{ik}\}$, — выходной символ ε . В соответствии с матрицей A осуществляется переход в следующее состояние. За T тактов источник выдаст последовательность

$$\varepsilon = \varepsilon_1, \dots, \varepsilon_T.$$

Вероятность ее выдачи задается формулой:

$$P(\varepsilon|A, B, \pi) = \sum_{Q^T} \pi_{q_1} \prod_{s=1}^T a_{q_{s-1}q_s} \cdot b_{q_s}(\varepsilon_s), \quad (1)$$

где Q^T — множество всех последовательностей состояний длины T , $a_{q_1q_2} \in A$, $b_q(\varepsilon_i) \in B$.

Из формулы (1) следует, что число арифметических операций, необходимых для подсчета функции правдоподобия, пропорционально N^T .

Под обучением вероятностного источника понимается адаптивное изменение матриц A , B и вектора π .

Зафиксируем последовательность $\varepsilon = \varepsilon_1, \dots, \varepsilon_T$. Алгоритм Баума обучения вероятностных источников [3, 4] использует оценку функции правдоподобия $P(\varepsilon|A, B, \pi)$. Для эффективного «счета» используется прямая и обратная функции частичного правдоподобия для $0 < t < T$:

$$\mu_t(i) = P(\varepsilon_1 \dots \varepsilon_t | q_t = i), \quad \nu_t(j) = P(\varepsilon_{t+1} \dots \varepsilon_T | q_t = j).$$

«Прямая» функция $\mu_t(i)$ — это условная вероятность того, что к моменту времени t выдано слово $\varepsilon_1 \dots \varepsilon_t$ и в этот же момент времени вероятностный источник находится в состоянии под номером i . «Обратная» функция $\nu_t(j)$ есть условная вероятность наблюдения слова $\varepsilon_{t+1} \dots \varepsilon_T$, начиная с момента времени $t + 1$, при условии, что источник находился в состоянии с номером j . Если инициализировать начальные значения введенных функций естественным образом:

$$\mu_1(i) = \pi_i b_i(\varepsilon_1), \quad \nu_T(j) = 1, \quad i, j = 1, \dots, |Q|,$$

то можно заметить, что прямая и обратная функции правдоподобия допускают рекурсивный счет:

$$\mu_{t+1}(i) = b_i(\varepsilon_{t+1}) \sum_{j=1}^{|Q|} \mu_t(j) a_{ji}, \quad \nu_t(i) = b_i(\varepsilon_{t+1}) \sum_{j=1}^{|Q|} \nu_{t+1}(j) a_{ij},$$

$$t = T - 1, \dots, 1, \quad 0 \leq i \leq |Q|.$$

Верно равенство:

$$P(\varepsilon|A, B, \pi) = \sum_{i=1}^{|Q|} \mu_T(i). \quad (2)$$

Вычисление вероятности по формуле (2) требует выполнения $O(|Q|T^2)$ операций, что для, например, $|Q| = 10$ и $|\varepsilon| = 5$, в 2000 раз меньше, чем для прямых вычислений по формуле (1). С помощью введенных функций можно обучать вероятностные источники чаще выдавать слово $\varepsilon = \varepsilon_1 \dots \varepsilon_T$. Если заменить матрицы A, B в вероятностном источнике на новые матрицы \hat{A}, \hat{B} , полученные по нижеприведенным формулам (3), то получим

$$P(\varepsilon|\hat{A}, \hat{B}) \geq P(\varepsilon|A, B)$$

и, более того, итеративное применение формул (3) гарантирует нам сходимость к оценкам максимального правдоподобия параметров A, B , то есть к локальному максимуму функции $f(A, B) = P(\varepsilon|A, B)$. Формулы пересчета параметров источника таковы:

$$\hat{a}_{ij} = \frac{\sum_{t=1}^T \mu_{t-1}(i) a_{ij} b_j(\varepsilon_t) \nu_t(j)}{\sum_{t=0}^{T-1} \mu_t(i) \nu_t(i)}, \quad \hat{b}_{jk} = \frac{\sum_{t=1, \varepsilon(t)=k}^T \mu_t(j) \nu_t(j)}{\sum_{t=1}^T \mu_t(j) \nu_t(j)}, \quad (3)$$

где $1 \leq i, j \leq |Q|, 1 \leq k \leq |E|$.

Распространенной задачей теории вероятностных источников является определение наиболее вероятной последовательности состояний источника, при условии выдачи им данного слова $\varepsilon = \varepsilon_1 \dots \varepsilon_T$. Эту задачу решает алгоритм Витерби, описанный на русском языке в [3]. Пусть $\psi_t(i)$ — начало длины $t-1$ наиболее вероятной частичной последовательности состояний длины t , оканчивающаяся в состоянии q_i . Пусть

$$\phi_t(i) = P(\psi_t(i)|A, B, \pi).$$

Первоначально положим $\phi_0(1) = 1, \phi_0(i) = 0$ и $\psi_0(1) = q_1$, где $2 \leq i \leq |Q|$. Тогда верна формула рекурсивного подсчета:

$$\varphi_t(j) = b_j(\varepsilon_t) \cdot \max_{1 \leq i \leq |Q|} \{\varphi_{t-1}(i) a_{ij}\}.$$

Следующее наиболее вероятное состояние $q_{i(t)}$ источника находится из $i(t) = \arg \max_{1 \leq j \leq |Q|} \{\varphi_t(j)\}$. Работа алгоритма заканчивается при $t = T$.

Использование вероятностных источников связано с рядом трудностей. Большинство известных методов распознавания, основанных на вероятностных источниках, предполагают подсчет функции правдоподобия $P(\varepsilon|A, B, \pi)$. Если число состояний источника велико, то подсчет этой функции занимает большое количество операций (а именно, $O(|Q|T^2)$, где $|Q|$ — число состояний источника, а T — длина слова).

Обучение источника требует присутствия «учителя», так как при длительном итеративном применении формул обучения (3), мы научим источник выдавать слова только из обучающей выборки. Проблема остановки обучения вероятностного источника до сих пор теоретически не решена. Обычно обучение останавливает «учитель».

Еще одна трудность — определение числа состояний источника, способного решать поставленную задачу.

Заметим, что вероятностный источник, согласно формуле (1), задает распределение вероятностей на словах длины T , и не всякое распределение может быть задано вероятностным источником. В самом деле, после обучения источника получается вероятностное распределение, к примеру, на булевом кубе E_2^N , $E_2 = \{0, 1\}$, то есть функция

$$f(x_1, \dots, x_N) : E_2^N \rightarrow [0, 1], \quad \sum_{\sigma} f(\sigma) = 1,$$

здесь суммирование ведется по всем вершинам σ булева куба.

Имеет место

Утверждение 1. *Для любого натурального k существует натуральное $N > k + 1$ и распределение $f(x_1, \dots, x_N) : E_2^N \rightarrow [0, 1]$, такое, что не существует вероятностного источника $L = (Q, E_2, A, B, \pi)$ с k состояниями, реализующего распределение f .*

Рассмотрим $N = k + 2$ и возьмем распределение следующего вида

$$f(1, 1, \dots, 1, 0) = 1, \quad f(a_1, a_2, \dots, 0, \dots, a_N) = 0$$

для всех i и всех $a_1, a_2, \dots, a_{i-1}, 0, a_{i+1}, \dots, a_N$. Заметим, что для любого состояния q источника L найдется $0 \leq i < k$, такое, что путь

$$q_j q_{i1} q_{i2} \dots q_{it}, \quad \text{где } q_{it} = q$$

из некоторого q_j в q , где $\pi_j > 0$ имеет ненулевую вероятность

$$a_{q_j q_{i1}} a_{q_{i1} q_{i2}} \dots a_{q_{it-1} q_{it}} > 0.$$

Поскольку $P(a_1 a_2 \dots a_i \dots a_N | L) = 0$ при $a_i = 0$, значит $b_q(0) = 0$ для всех q достижимых за число шагов не превосходящее k . А это означает, что $b_q(0) = 0$ для всех достижимых состояний источника L , и равенство $P(11 \dots 10 | L) = 1$ невозможно.

Систему вероятностных источников $\{L_i\}$ используют для распознавания через синтез. В самом деле, для каждого слова ε по формуле (1) мы знаем вероятность $P(\varepsilon | L_i)$ того, что его выдал данный вероятностный источник. Выберем в качестве результата распознавания то i , у которого $P(\varepsilon | L_i)$ максимально.

2. Примеры применения вероятностных источников

А теперь рассмотрим примеры применения вероятностных источников в современных инженерных задачах. Широко распространены вероятностные источники в составе скрытых марковских моделей в распознавании речи.

Вероятностные источники моделируют график звукового давления при произнесении фонем. Работа по обучению вероятностных источников-фонем происходит с учителем при использовании реальных записей фонем. Объемы реальных записей в некоторых системах распознавания речи достигают нескольких гигабайтов. Качество распознавания фонем вероятностными источниками превзошло уровень человеческого распознавания фонем. Вероятностный источник фонемы всегда имеет «монотонную» матрицу переходных вероятностей. То есть матрица A имеет верхнетреугольный вид.

Переход от распознавания фонем к распознаванию слов реализуется объединением вероятностных источников-фонем, составляющих

слово. Для каждого слова мы можем автоматически построить его вероятностный источник, как показано на рис. 1. Последнее состояние источника-фонемы соединяется с начальными состояниями источников фонем — возможных продолжений этой фонемы.

Начальные состояния на рисунке закрашены штрихом, поглощающие (соответствующие тишине) выделены цветом.

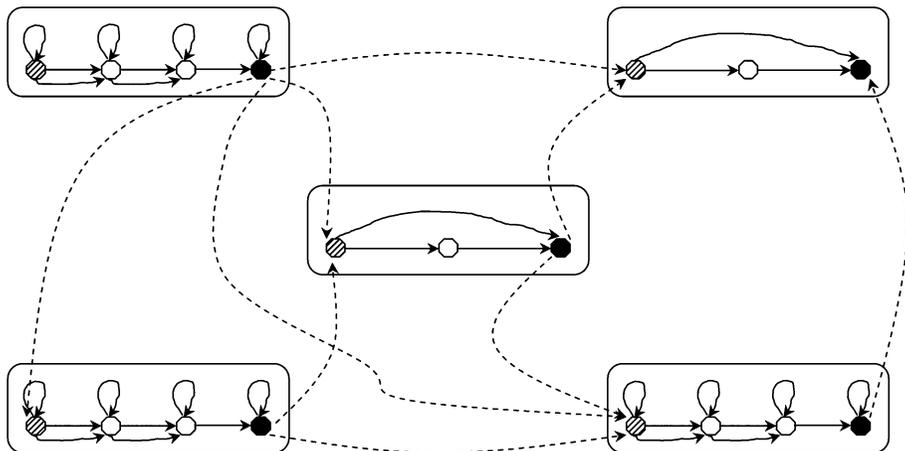


Рис. 1. Пример вероятностного источника для уровня слов.

Для распознавания неизвестного слова, составленного из цепочки фонем, отыскиваем по алгоритму Витерби наиболее вероятную последовательность фонем во вновь сформированном источнике, которая и будет являться результатом распознавания. Проблемы и перспективы данной методики в распознавании речи представлены в [5].

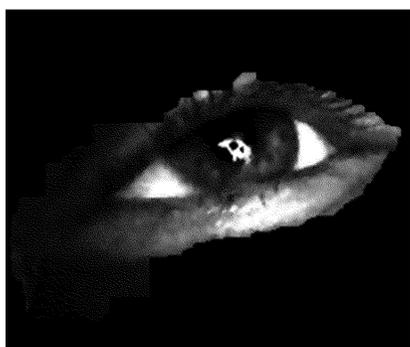
Для успешного хранения и использования большого объема данных необходима их автоматическая классификация. Например, имея информацию о том, содержит ли данный сигнал в себе только речь, или музыку, или звуки природы, мы можем кодировать его разными кодеками, существенно увеличивая эффективность кодирования.

Поскольку основную массу сигналов составляет именно музыка и речь, то первой и основной задачей является классификация сигнала на музыку/речь. В статье [2] можно найти подробное описание классификатора музыка/речь. В этой работе вероятностные источники

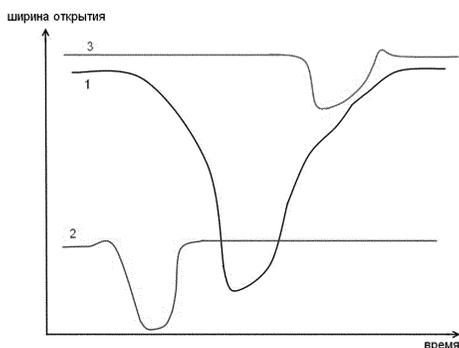
используются для сглаживания результатов распознавания. Обучаются четыре вероятностных источника, соответственно для моделирования музыки, речи, не-музыки и не-речи. И на основе функции правдоподобия (с какой вероятностью каждый источник выдал бы заданный сигнал) принимается решение о классификации сигнала. В качестве материала для обучения выступают заранее известные характеристики сигнала, например кепстральные коэффициенты, число пересечений нуля сигналом и т. д. Обучение источников в работе [2] проводилось на десятках часов аудио-записей, точность распознавания сигнала оказалась выше 97%.

Аппарат вероятностных источников и скрытых марковских моделей также используется при распознавании рукописных текстов и предсказании погоды. В задачах метеорологии вероятностные источники вновь и вновь обучаются на поступающих несколько раз в день данных с метеостанций. Прогноз погоды производится на основе предсказания наиболее вероятного состояния, например, с помощью алгоритма Витерби.

Автор применил данную методику к задаче распознавания засыпаний человека. Задача важна для определения засыпания водителей транспортных средств и авиадиспетчеров. Видеокамера снимала область, содержащую глаза испытуемого.



а)



б)

Рис. 2. а) Черно-белое изображение глаза человека. б) Возможные виды морганий. 1 — человек смотрит в камеру. 2 — человек смотрит вниз или вбок. 3 — избыток света.

Полученное изображение обесцвечивалось и вычислялась ширина открытия глаза в данный момент времени¹. Таким образом, получалась последовательность чисел — степеней открытия глаза. По этой квантованной последовательности и надо было предсказать засыпание водителя. Связь длительности и частоты морганий с состоянием бодрствования человека описана в [7].

Для предсказания засыпания требуется точное определение числа и длительности морганий. Несмотря на кажущуюся простоту, это — непростая задача. Дело в том, что человек может по-разному держать голову во время вождения, и жестко закрепленные видеокамеры будут фиксировать различного вида сигналы, которые могут быть неправильно классифицированы. Например, если человек смотрит из-подлобья или отвел взгляд в сторону, то обычный алгоритм определения моргания по порогу — если график упал ниже порога, то фиксируем моргание — не сработает или сработает неверным образом. Также ошибки возможны при резкой смене освещения.

Виды графиков зависимости определяемой ширины открытия глаза в зависимости от времени, в этих случаях, представлены на рис. 2б. Все это говорит о том, что необходимо иметь устройство, распознающее общий вид графика моргания, учитывающее все его характеристики целиком. В качестве этого устройства можно использовать вероятностный источник, обученный на типичных морганиях, полученных в результате реальных экспериментов. То есть, имея графики с рисунка 2б, предлагается обучить вероятностный источник на рис. 3.

Такой выбор источника неслучаен. Число и связь состояний отражают механику моргания: 1-ое состояние — глаз открыт, 2-ое — глаз полуоткрыт, 3-е — глаз полностью закрыт, 4-ое — глаз полузакрыт, и, наконец, 5-ое — глаз снова открыт. Все возможные «стадии» описаны.

Глаз может находиться в определенном состоянии довольно долго, поэтому обязательны переходы состояний в себя. Как показали испытания, при скорости камеры 30 кадров в секунду глаз успевает

¹За ширину открытия глаза было положено количество белого цвета на изображении.

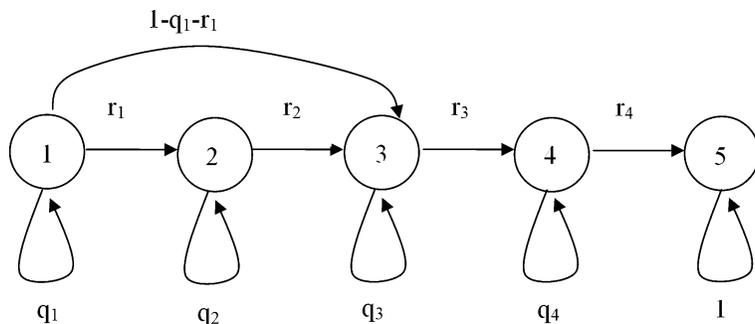


Рис. 3. Вероятностный источник для определения морганий. На рисунке изображены переходы только с ненулевой вероятностью.

закрыться, минуя стадию полузакрытого глаза, чему соответствует наличие ненулевого перехода из 1-го состояния в 3-е. Однако, веко не может подняться достаточно быстро, чтобы миновать стадию «полукоткрытого» глаза — этому соответствует отсутствие перехода между 3-им и 5-ым состоянием². Совершенно аналогично, мы можем обучить другой вероятностный источник L выдавать «не моргания». Обучив описанный выше источник для морганий K на разного рода морганиях, мы получим решающее правило: является ли поданная на вход последовательность чисел $\varepsilon_1, \dots, \varepsilon_N$ морганием будет определяться исходя из того, какая вероятность больше:

$$P(\varepsilon_1, \dots, \varepsilon_N | K) \text{ или } P(\varepsilon_1, \dots, \varepsilon_N | L)$$

При этом длительность моргания определяется исходя из наиболее вероятной последовательности состояний q_1, \dots, q_N для данного источника с помощью вышеописанного алгоритма Витерби, где N — длина входной последовательности. Этот метод был реализован в программе и протестирован. Результат предсказания засыпаний — свыше 98% на качественных видеозаписях, что существенно превосходит пороговый метод определения морганий. Более того, алгоритм обнаруживает моргания, пропущенные экспертом. Подробнее эта модель описана в [6]. На основании статьи [7] о связи динамики мор-

²Точнее, переход, конечно, есть, но ему приписана нулевая вероятность.

ганий с засыпаниями, можно проверить возможность предсказания засыпания с помощью аппарата вероятностных источников.

3. Заключение

Из вышеприведенного следует, что обучение вероятностного источника — задача, зависящая от числа состояний источника. Это объясняет вычислительную сложность задачи распознавания через синтез. Так как решающее правило здесь — функция правдоподобия, необходимая для классификации и распознавания, напрямую зависит от числа состояний источника. Практика показывает, что для того, чтобы обучить источник «большому множеству», требуется большое число состояний, и, как следствие, вероятностные источники могут применяться и применяются либо в задачах, где не требуется вычислений в реальном времени, либо для моделирования относительно простых процессов.

Успех применения метода скрытых марковских моделей обычно связан с тем, что автор хорошо знает и может декомпозировать моделируемый объект. Метеорологи в задаче прогнозирования знают классы состояний погоды, филологи и графологи четко выделяют отличимые структурные единицы речи и почерка. Так же и автор статьи представил моргание в виде последовательных состояний глаза человека, что и послужило предпосылкой для создания моделирующего источника определенного вида, что, в итоге, и привело к успеху модели.

Список литературы

- [1] Кудрявцев В. Б., Алешин С. В., Подколзин А. С. Введение в теорию автоматов. М.: Наука, 1985.
- [2] Бухараев Р. Г. Введение в теорию вероятностных автоматов. М.: Наука, 1985.
- [3] Dempster A. P., Laird N. M., Rubin D. B. Maximum likelihood from incomplete data via the EM algorithm // J. Royal Stat. Soc. Vol. 39. 1977. P. 1–87.

- [4] Music and speech detection system based on Hidden Markov Models and Gaussian Mixture Models / White paper by compure.com.
- [5] Левинсон С. Е. Структурные методы автоматического распознавания речи // ТИИЭР. № 11. Ноябрь 1985. С. 100–126.
- [6] Baum L. E. An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes // Inequalities. Vol. 3. 1972. P. 1–8.
- [7] Бабин Д. Н., Мазуренко И. Л., Холоденко А. Б. О перспективах создания системы автоматического распознавания слитной устной русской речи // Интеллектуальные системы. Т. 8, вып. 1–4. 2004. С. 45–70.
- [8] Пархоменко Д. В. Применение марковских моделей к проблеме распознавания засыпаний // Вопросы атомной науки и техники. Вып. 4. 2008. С. 54–60.
- [9] Electrooculogram Analysis and Development of a System for Defining Stages of Drowsiness / Report from Linköping University, Dept. Biomedical Engineering. LiU-IMT-EX-351.

