УДК 004.934

ФУНКЦИЯ ПЛОТНОСТИ ДЛИТЕЛЬНОСТИ СОСТОЯНИЙ СММ. ПРЕИМУЩЕСТВА И НЕДОСТАТКИ

Балакшин П.В.

Санкт-Петербургский государственный университет информационных технологий, механики и оптики, Санкт-Петербург, e-mail: pvbalakshin@gmai.com

В статье кратко раскрыты понятия скрытых марковских моделей (СММ) и наблюдаемых марковских моделей. Рассмотрен вопрос о способах задания плотности длительности состояний. Произведены расчет и сравнение трудоемкости ряда алгоритмов распознавания речи при различном задании плотности длительности состояний. Определены преимущества и недостатки использования и управления функцией плотности длительности состояний СММ.

Ключевые слова: распознавание речи, скрытые марковские модели, длительность состояний, функция плотности длительности состояний

DENSITY FUNCTION OF DURATION IN STATES OF HMM. ADVANTAGES AND LIMITATIONS

Balakshin P.V.

Saint-Petersburg State University of Information Technologies, Mechanics and Optics, Saint-Petersburg, e-mail: pvbalakshin@gmai.com

This article briefly disclosed the concept of hidden markov models (HMM) and observable markov models. A question about the ways of density function of duration in states assignment was investigated. Calculation and comparison of some speech recognition algorithms laboriousness with different ways of density function of duration in states assignment were made. Advantages and limitations of use and control of density function of duration in states of HMM were detected.

Keywords: speech recognition, hidden markov models, duration in states, density function of duration in states

Процесс взаимодействия человека с ЭВМ стал одним из наиболее важных вопросов развития компьютерной техники с момента появления самих ЭВМ. Сначала технологи «общались» с ней через посредника-программиста. Затем был диалоговый интерфейс, после графический. Но человечество всегда искало более простое, удобное, естественное решение.

Поэтому нетрудно понять, что голосовой интерфейс – это тема, которая на протяжении последних пятидесяти лет привлекает внимание ученых и инженеров всего мира. Голосовой интерфейс на языке пользователя – это наилучшее решение. Действительно, ведь речь – это наиболее натуральная, удобная, эффективная и экономичная форма человеческого взаимодействия. Первые конкретные шаги в данной области были предприняты в 1971 году по заказу министерства обороны США [1].

В связи с быстрым и постоянным развитием как науки, так и самой жизни человека, понятия скорость и время становятся все более значимыми. Бурный рост информационных технологий позволил несколько снизить время, затрачиваемое на передачу и обработку информации. И на первый план стали выходить понятия точности, достоверности и надежности этой информации.

В системах автоматического распознавания речи важнейшим показателем является точность распознавания слитной речи, которая определяется отношением количества верно распознанных слов к сумме всех распознаваемых слов, пропущенных слов и лишних (неверно распознанных) слов [5]. В настоящее время, этот показатель стремиться к значению 95 и выше процентов.

Повышение указанного показателя является важной и актуальной задачей. Подтверждением этого могут служить:

1) множество научно-исследовательских центров, объединенных в Международную Ассоциации Речевого Взаимодействия (International Speech Communication Association);

2) научно-технические конференции, крупнейшей из которых является ежегодная Interspeech;

3) различные программно-технические разработки (Dragon Naturally Speaking, IBM Via Voice, встроенное речевое управление в ОС Vista).

Следует выделить 2 аспекта, в которых ведутся исследования, связанные с повышением точности распознавания: аппаратный и программный. В данной работе нами уделено внимание лишь алгоритмам распознавания, которые являются частью программного аспекта. Все существующие ныне основные алгоритмы могут быть группированы следующим образом:

1. Распознавание на основе СММ – скрытых марковских моделей.

2. Распознавание на основе нейросети.

3. Гибридные модели.

Каждая из них имеет свои достоинства и недостатки. Прежде всего, это обусловлено конкретными задачами и местом применения. Но большинство исследователей отдают предпочтение скрытым марковским моделям. Таким образом, задачей исследования является модификация алгоритмов распознавания речи, основанных на СММ, для повышения точности распознавания. Базовые принципы алгоритмов распознавания речи были сформулированы начале 80-х годов прошлого столетия Л. Рабинером и Р. Шафером в книге «Цифровая обработка речевых сигналов» [1]. А в качестве опубликованных исследований о трудоемкости алгоритмов распознавания речи можно выделить только статью сотрудников компании SPIRIT из журнала «Цифровая обработка сигналов» [2].

Наблюдаемая марковская модель. Рассмотрим систему, которая в произвольный момент времени может находиться в одном из N различных состояний $S = s_1, s_2, s_3, \ldots, s_N$. В дискретные моменты времени система претерпевает изменение состояния (возможно, переходя при этом опять в тоже состояние) в соответствии с некоторым вероятностным правилом, связанным с текущим состоянием. Моменты времени, в которые происходит изменение состояния системы, будем обозначать через $t = 1, 2, ..., T$, а ее состояние в момент времени t через q_t. Полное вероятностное описание такой системы будет, вообще говоря, требовать задания как текущего состояния (в момент времени t), так и всех предыдущих. Для частного случая дискретной цепи Маркова первого порядка это вероятное описание требует знания только текущего и предыдущего состояний, т.е. сводится к виду [6]:

$$P\left\langle q_t = s_j \,\middle|\, q_{t-1} = s_i, q_{t-2} = s_k, \ldots \right\rangle = P\left\langle q_t = s_j \,\middle|\, q_{t-1} = s_i \right\rangle. \quad (1)$$

В дальнейшем мы будем рассматривать такие процессы, для которых правая часть в (1) не зависит от времени. В этом случае переходные вероятности a_{ij} определяются выражением:

$$a_{ij} = P\left\langle q_t = s_j \,\middle|\, q_{t-1} = s_i \right\rangle, \quad (2)$$

где $1 \le i, j \le N$ и обладают следующими свойствами: $a_{ij} \ge 0$, $\sum_{j=1}^{N} a_{ij} = 1$, поскольку удовлетворяет обычным вероятностным ограничениям.

Описанный выше стохастический процесс может быть назван наблюдаемой марковской моделью, так как выходом такого процесса в каждый момент времени является очередное состояние модели, которое соответствует физическому (наблюдаемому) событию.

Поясним описанную конструкцию на примере. Рассмотрим СММ, обладающую тремя состояниями и предназначенную для моделирования погоды. Предполагается, что раз в день (например, в полдень) состояние погоды описывается одной (и только одной) из следующих характеристик:

(1) состояние 1: осадки;
(2) состояние 2: облачно;
(3) состояние 3: ясно.

Матрица A, составленная из вероятностей перехода между состояниями, имеет вид:

$$A = \left\{a_{ij}\right\} = \begin{bmatrix} 0{,}4 & 0{,}3 & 0{,}3 \\ 0{,}2 & 0{,}6 & 0{,}2 \\ 0{,}1 & 0{,}1 & 0{,}8 \end{bmatrix} \quad (3)$$

Пусть известно, что день 1-й ($t = 1$) – ясный (т.е. имеем состояние 3). Можно задать следующий вопрос: какова вероятность (в соответствии с заданной моделью) того, что в последующие 5 дней последовательность состояний погоды будет иметь вид: «ясно-облачно-осадки-осадки-ясно-ясно»?

Таким образом, задана последовательность наблюдений $O = \left\{s_3, s_2, s_1, s_1, s_3, s_3\right\}$, соответствующая моментам времени $t = 1, 2, ..., 6$, и нужно определить вероятность появления этой последовательности для данной модели. Используя формулу Байеса и формулу (2) эта вероятность может быть записана и вычислена с помощью выражения:

$$\begin{aligned} P\left\langle O \,\middle|\, \text{модель} \right\rangle &= P\left\langle s_3, s_2, s_1, s_1, s_3, s_3 \,\middle|\, \text{модель} \right\rangle = \\ &= P\left\langle s_3 \right\rangle \cdot P\left\langle s_2 \,\middle|\, s_3 \right\rangle \cdot P\left\langle s_1 \,\middle|\, s_2 \right\rangle \cdot P\left\langle s_1 \,\middle|\, s_1 \right\rangle \cdot P\left\langle s_3 \,\middle|\, s_1 \right\rangle \cdot P\left\langle s_3 \,\middle|\, s_3 \right\rangle = \\ &= \pi_3 \cdot a_{32} \cdot a_{21} \cdot a_{11} \cdot a_{13} \cdot a_{33}. \end{aligned} \quad (4)$$

Т.е. $P\left\langle O \,\middle|\, \text{модель} \right\rangle = 1 \cdot 0{,}1 \cdot 0{,}2 \cdot 0{,}4 \cdot 0{,}3 \cdot 0{,}8 = 1{,}92 \cdot 10^{-3}$, где $\pi_3 = P\left\langle q_1 = s_3 \right\rangle$ – вероятность того, что начальное состояние 3 (т.е. ясно).

Плотность длительности состояний. Другой не менее важный и интересный вопрос звучит следующим образом: пусть модель находится в некотором известном состоянии. Какова вероятность того, что она останется в этом состоянии ровно d моментов времени? Эта вероятность может быть вычислена как вероятность последовательности наблюдений

$$O = \left\{ \underset{1}{s_i}, \underset{2}{s_i}, \underset{3}{s_i}, ..., \underset{d}{s_i}, \underset{d+1}{s_j} \neq s_i \right\},$$

из которой, с учетом введенной модели (т.е. формулы (4)) получаем:

$$P\left\langle O \middle| \text{модель}, q_1 = s_i \right\rangle = \left(a_{ii}\right)^{d-1} \cdot \left(1 - a_{ii}\right) = \rho_i\left(d\right) \quad (5)$$

Величина $\rho_i(d)$ есть не что иное, как дискретная функция плотности вероятности пребывания в течение времени d в состоянии i. Эта функция имеет вид показательной и представляет собой характеристику длительности данного состояния в СММ. Используя $\rho_i(d)$, нетрудно вычислить ожидаемое число повторений одного и того же состояния (то есть математическое ожидание времени непрерывного пребывания в этом состоянии):

$$\overline{d_i} = \sum_{d=1}^{\infty} d\rho_i\left(d\right) = \sum_{d=1}^{\infty} d\left(a_{ii}\right)^{d-1} \cdot \left(1 - a_{ii}\right) = \frac{1}{1 - a_{ii}}. \quad (6)$$

Так, согласно принятой модели, наиболее вероятное число следующих друг за другом ясных дней равно $\frac{1}{1-0{,}8} = \frac{1}{0{,}2} = 5$, облачных – 2,5 и осадками – 1,67.

Возможно, основной недостаток обычных СММ обусловлен используемым в них способом моделирования длительности состояний [6]. Как было показано (5), плотность вероятности $\rho_i(d)$ пребывания в состоянии s_i с переходной вероятностью a_{ii} имеет вид показательной функции. Однако для большинства физических сигналов такая плотность вероятности для длительности пребывания в состоянии (т.е. плотность длительности состояния) является неприемлемой. Вместо этого длительность состояния многие исследователи считают нужным моделировать явно, в том или ином аналитическом виде. Рис. 1 и рис. 2 иллюстрируют (для одной из пар состояний модели s_i и s_j) различия между СММ, с явно и неявно заданными функциями плотности длительности состояний.

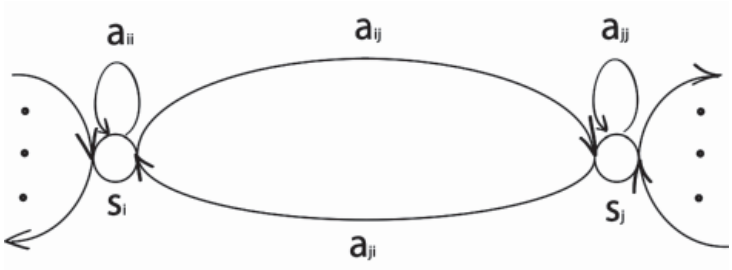

Рис. 1. СММ, в которой функции плотности длительности состояний имеют вид прерывистых показательных

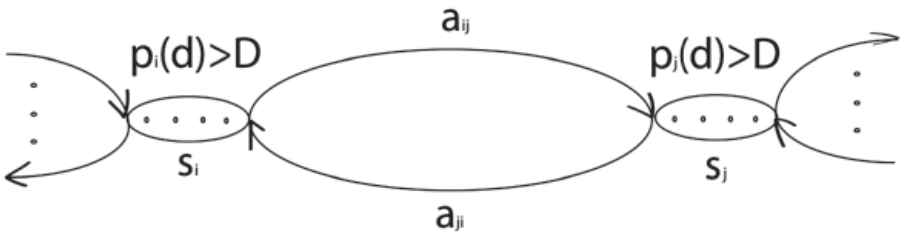

Рис. 2. СММ, в которой функции плотности длительности состояний заданы явно

На рис. 1 эти состояния имеют плотности длительности вида (5), полученные соответственно по переходным вероятностям a_{ii} и a_{jj}. На рис. 2 коэффициенты таких вероятностей «перехода в себя» установлены равными 0, и заданы явные функции плотности длительности состояний. В этом случае переход из некоторого состояния осу-

ществляется только после того, как будет порождено столько наблюдений, сколько их было определено в соответствии с функцией плотности длительности этого состояния. Число этих переходов задано как D.

Целесообразность введения явного описания для плотностей длительностей состояний подтверждается тем, что для многих задач это значительно улучшает качество моделирования, а, значит, и качество распознавания. Получается, что мы некоторым образом нормируем каждое состояние. Подобные подходы уже не раз показывали свою пригодность. Так в 1960 году Динез и Мэтьюз ввели концепцию временной нормализации [1].

Однако подобной модели присущи и недостатки. Одним из них является значительный рост вычислений, обусловленный введением изменяющихся длительностей состояний, что следует непосредственно из определения и условий инициализации переменной $\alpha_t(j)$ для алгоритмов прямого-обратного хода и Витерби [1], который в последнее время выделяется практически как самостоятельный метод динамического программирования [4]. При этом объем требуемой памяти увеличивается примерно в D раз, а число необходимых вычислительных операций – примерно в $\frac{D^2}{2}$ раз по сравнению с обычными СММ.

Так, при количестве отсчетов времени $T = 200$ и числе состояний $N = 7$ для вычисления всех $\alpha_t(j)$, где $1 \le j \le N$, алгоритм прямого-обратного хода потребует порядка

$$N^2 \cdot T = 49 \cdot 200 = 9800 \approx 10^4 \text{ операций.}$$

Так, для значения D порядка 25 (что приемлемо для многих задач, связанных с обработкой речи) объем вычислений увеличивается примерно в 300 раз:

$$N^2 \cdot T \cdot \frac{D^2}{2} = 9800 \cdot \frac{625}{2} \approx 3 \cdot 10^6 \text{ операций}$$

Однако это по-прежнему меньше, чем при использовании прямых вычислений [1], требующих

$$2 \cdot T \cdot N^T \approx 4 \cdot 10^{171} \text{ операций.}$$

Другая проблема, связанная с моделями этого типа, состоит в том, что помимо обычных параметров СММ для них необходимо оценивать также большое (равное D) число новых параметров, связанных с каждым состоянием. Кроме того, при фиксированном числе наблюдений T обучающее множество с явно заданной плотностью длительности состояний содержит, в среднем, меньше переходов между состояниями и, следовательно, меньше данных для оценивания величин $\rho_i(d)$, чем это было бы в случае стандартной СММ.

Но в тоже время функция плотности длительности состояний позволяет модифицировать алгоритм Витерби, максимизируя не предыдущее состояние, а длительность текущего.

Заключение. Таким образом, можно сформулировать следующие преимущества и недостатки использования и управления функцией плотности длительности состояний скрытых марковских моделей. Среди преимуществ стоит выделить:

(1) Значительно улучшается качество моделирования, а, значит, и качество распознавания.

(2) Возможность использования для модификации алгоритма Витерби, позволяющего:

(a) существенно быстрее восстанавливать порядок состояний модели благодаря уже имеющейся информации о длительности каждого состояния;

(b) автоматически разрешать равновероятностные переходы.

Среди недостатков отметим:

(1) Значительный рост вычислений, обусловленный введением изменяющихся длительностей состояний.

(2) Увеличение объема требуемой памяти.

(3) Необходимость оценивать большое число новых параметров, связанных с каждым состоянием (примерно равное D).

Нетрудно видеть, что некоторые преимущества и недостатки почти противоречат друг другу. Данная ситуация легко объясняется тем, что рост вычислений приходится более короткий промежуток времени, что в итоге должно дать преимущество в точности и, вероятно, скорости распознавания.

Данная работа была выполнена автором в качестве дополнительной проверки рациональности реализации разработанного модифицированного алгоритма Витерби.

Работа выполнена при поддержке гранта Правительства Петербурга № 3.11/04-06/50.

Список литературы

1. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов: пер. с англ.; под. ред. М.В. Назарова и Ю.Н. Прохорова. – М.: Радио и связь, 1981. – 496 с.

2. Иконин С.Ю., Сарана Д.В. Система автоматического распознавания речи SPIRIT ASR Engine // Цифровая обработка сигналов. – 2003. № 4.

3. Гультяева Т.А. Скрытые Марковские процессы. – Новосибирск: Изд-во НГТУ.

4. Аграновский А.В., Леднов В.А. Теоретические аспекты обработки и классификации речевых сигналов. – М.: Радио и связь, 2004. – 164 с.

5. Tebelskis J. Speech Recognition using Neural Networks. School of Computer Science Carnegie Mellon University. – 1995. – 190 p.

6. Rabiner L.R. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition // IEEE. – 1989. – Vol. 77, №2. – C. 257–286.