УДК 004.855.5

Д.А. Гефке, П.М. Зацепин

Применение скрытых марковских моделей для распознавания звуковых последовательностей

D.A. Gefke, P.M. Zatsepin

The Application of Hidden Markov Models for the Acoustic Wave Recognition

В работе рассмотрен аппарат скрытых марковских моделей применительно к решению задачи распознавания звуковых последовательностей (задача распознавания раздельной речи). Реализована и протестирована экспериментальная система, основанная на использовании кепстральных коэффициентов (MFCC) и коэффициентов линейного предсказания (LPC). Получены практические оценки точности распознавания для различных параметров модели.

Ключевые слова: скрытая марковская модель, кепстральные коэффициенты (MFCC), коэффициенты линейного предсказания (LPC).

The paper examines the application of hidden Markov models for acoustic wave recognition (isolated word recognition problem). The experimental system using MFCC and LPC coefficients was developed and tested. The estimates of recognition probabilities for various system parameters were obtained.

Key words: hidden Markov model, MFCC coefficients, LPC coefficients.

Введение. Математический аппарат скрытых марковских моделей (СММ) представляет собой универсальный инструмент описания стохастических процессов, для работы с которыми не существует точных математических моделей, а их свойства меняются с течением времени в соответствии с некоторыми статистическими законами. Наиболее широкое применение СММ нашли при решении таких задач, как распознавание раздельной и слитной речи, анализ изображений, видео, последовательностей ДНК и ряда других [1].

В данной работе СММ рассматриваются применительно к решению задачи распознавания звуковых последовательностей (задача распознавания раздельной речи). В открытой литературе присутствует достаточное количество общей теоретической информации относительно аппарата СММ. Однако практические аспекты применения СММ, в частности, выбор параметрического вектора, числа состояний, параметров гауссовых смесей, оценки точности распознавания и т.д., освещены мало.

Цель данной работы – реализация аппарата скрытых марковских моделей и проведение ряда экспериментов для поиска оптимальных параметров модели по критерию уменьшения ошибки обобщения (максимизация вероятности распознавания «своих» образцов и минимизация вероятности распознавания ложных образцов) применительно к решению задачи распознавания раздельной речи (*isolated word recognition problem*).

Описание скрытой марковской модели. В основе скрытой марковской модели лежит конечный автомат, состоящий из N-состояний, называемых *скрытыми*. Переходы между состояниями в каждый дискретный момент времени t не являются детерминированными, а происходят в соответствии с вероятностным законом и описываются матрицей вероятностей переходов A_{NN}. Схематическое изображение диаграммы переходов между состояниями СММ приведено на рисунке 1.

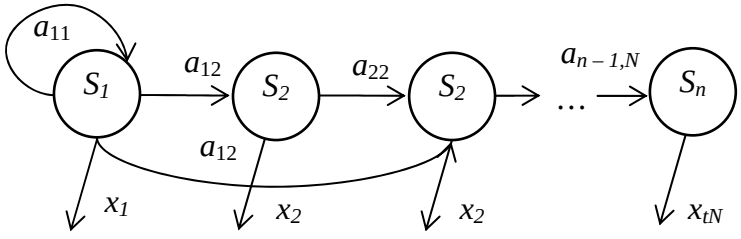


Рис. 1. Структурная схема переходов в СММ

Нахождение модели в некотором состоянии i соответствует определенной стационарности наблюдаемого сигнала на ограниченном временном интервале. Появляется простая физическая интерпретация СММ: рассматривается процесс, который иногда скачкообразно меняет свои характеристики [1].

При осуществлении очередного перехода в новое состояние i в момент времени t происходит генерация выходного вектора x_t, называемого *параметрическим вектором*, в соответствии с многомерной функцией распределения вероятностей $f_j(x)$. Результатом работы скрытой марковской модели является последовательность векторов (*наблюдений*) $[(\{x\})]_1, x_2, ..., x_T\}$ длиной T. Достоинством СММ является возможность обработки последовательностей и сигналов разной длины, что затруднено при работе с искусственными нейронными сетями, в частности [1].

Функция плотности вероятностей $f_j(x)$ для состояния j описывается, как правило, взвешенной гауссовой смесью:

$$f_j(x) = \sum_{i=1}^{M} w_i p_i(x), \qquad (1)$$

где M – количество компонент смеси; w_i – вес компонента смеси; $p_i(x)$ – нормальное распределение для D-мерного случая.

Функция $p_i(x)$ описывается следующим выражением:

$$p_i(x) = \frac{1}{2\pi^{\frac{D}{2}} |\sigma_i|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2(x-\mu_i)^T \sigma_i^{-1} (x-\mu_i)} \right\}, \qquad (2)$$

где D – размерность вектора; μ_i – вектор математического ожидания; σ_i – матрица ковариации.

Работа со скрытыми марковскими моделями, как и с любой другой адаптивной экспертной системой, осуществляется в два этапа [2]:

1) обучение – определение параметров модели – алгоритм Баума-Велча (*forward-backward, Baum-Welch re-estemation*);

2) определение – какова вероятность того, что наблюдаемая последовательность векторов $[(\{x\})]_1, x_2, ..., x_T\}$ была сгенерирована данной моделью – алгоритм максимума правдоподобия (Витерби).

Далее приводится краткое описание вышеперечисленных алгоритмов.

3) обучение скрытой марковской модели.

Процесс обучения скрытой марковской модели заключается в определении с помощью набора обучающих образцов следующих параметров:

– матрицы вероятностей переходов между состояниями A_{NN};

– параметров гауссовых смесей (математическое ожидание, матрица ковариации и веса) для каждого состояния.

Для решения этих задач совместно применяются два итерационных алгоритма: forward-backward и Baum-Welch re-estemation.

В алгоритме forward-backward вводятся две функции: прямого распространения вероятности $a_j(t)$ и обратного $\beta_j(t)$.

Значение величины $a_j(t)$ представляет собой вероятность наблюдения последовательности векторов $[(\{x\})]_1, x_2, ..., x_t\}$ и нахождения СММ в состоянии j в момент времени t:

$$a_j(t) = P(x_1, x_2, ..., x_t | state_t = j). \qquad (3)$$

Величины $a_j(t)$ и $a_j(t-1)$ связаны итерационным выражением:

$$a_j(t) = \left[\sum_{i=2}^{N-1} a_i(t-1) A_{ij} \right] f_j(x_t), \qquad (4)$$

где A_{ij} – вероятность перехода из состояния i в состояние j; $f_j(x_t)$ – вероятность наблюдения вектора x_t в состоянии j.

Обратная функция $\beta_j(t)$ представляет собой вероятность нахождения СММ в состоянии j в момент времени t с последующим наблюдением последовательности $[(\{x\})]_{t+1}, x_{t+2}, ..., x_T\}$:

$$\beta_j(t) = P(x_{t+1}, x_{t+2}, ..., x_T | state_t = j).$$

Величины $\beta_j(t)$ и $\beta_j(t+1)$ связаны аналогичным образом:

$$\beta_j(t) = \sum_{i=2}^{N-1} A_{ji} f_i(x_{t+1}) \beta_i(t+1). \qquad (5)$$

Величины $a_j(t)$ и $\beta_j(t)$ позволяют определить вероятность нахождения СММ в состоянии j в момент времени t при наблюдении последовательности $[(\{x\})]_1, x_2, ..., x_t\}$:

$$L_j(t) = \frac{1}{P} a_j(t) \beta_j(t), \qquad (6)$$

где $P = a_N(T)$ – общая вероятность наблюдения последовательности $[(\{x\})]_1, x_2, ..., x_t\}$ данной СММ.

Алгоритм Баума-Велча (Baum-Welch re-estemation) на очередном шаге обучения позволяет, используя вышеприведенные выражения, сделать переоценку параметров модели [2].

Пусть имеется R обучающих образцов, тогда вероятность перехода из состояния i в состояние j определяется как:

$$\tilde{A}_{ij} = \frac{\sum_r^R = 1 \frac{1}{P_r} \sum_{t=1}^{T_{r-1}} a_i^r(t) A_{ij} f_j(x_{t+1}^r) \beta_j^r(t+1)}{\sum_r^R = 1 \frac{1}{P_r} \sum_{t=1}^{T_r} a_i^r(t) \beta_j^r(t)}. \qquad (7)$$

Для каждого состояния j и для каждой компоненты гауссовой смеси m математическое ожидание, матрица ковариации и вес определяются следующими выражениями:

$$\hat{\mu}_{jm} = \frac{\sum_{r=1}^{R}\sum_{t=1}^{T_r} L_{jm}^{r}(t)x_t^r}{\sum_{r=1}^{R}\sum_{t=1}^{T_r} L_{jm}^{r}(t)};$$

$$\hat{\sigma}_{jm} = \frac{\sum_{r=1}^{R}\sum_{t=1}^{T_r} L_{jm}^{r}(t)\left[(x)\right]_t^r - \hat{\mu}_{jm}^r\left(\left[(x)\right]_t^r - \hat{\mu}_{jm}^r\right)^T}{\sum_{r=1}^{R}\sum_{t=1}^{T_r} L_{jm}^{r}(t)}; \quad (8)$$

$$\hat{w}_{jm} = \frac{\sum_{r=1}^{R}\sum_{t=1}^{T_r} L_{jm}^{r}(t)}{\sum_{r=1}^{R}\sum_{t=1}^{T_r} L_{jm}^{r}(t)}.$$

Для качественного обучения скрытой марковской модели требуется множество образцов сигнала: от нескольких десятков до нескольких сотен экземпляров. Также необходимо соблюдать условие линейной независимости обучающих образцов, в противном случае, в процессе обучения происходит вырождение матрицы ковариации, следствием чего является полная неработоспособность модели [2].

При практической работе со скрытыми марковскими моделями приходится решать ряд ключевых задач:

1) выбор системы параметрических векторов, например, для распознавания речи используются кепстральные коэффициенты (MFCC), коэффициенты линейного предсказания (LPC) и ряд других;

2) разработка алгоритма нормализации параметрических векторов;

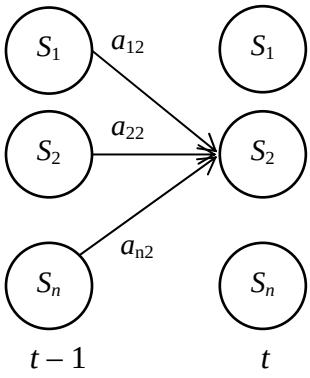
3) выбор количества состояний модели N и числа компонент гауссовой смеси M;

4) первоначальная сегментация обучающих векторов для нахождения приближенных значений математических ожиданий гауссовых смесей на первоначальном шаге обучения и т.д.

Необходимо заметить, что нет универсального алгоритма определения вышеперечисленных параметров и в каждом конкретном случае, в зависимости от решаемой задачи, может потребоваться проведение огромного количества экспериментов, прежде чем будут достигнуты требуемые результаты точности распознавания [2].

В процессе обучения может возникнуть ситуация, когда значения вероятностей в знаменателе вышеприведенных выражений будут иметь очень маленькие значения (близкие к нулю), что приведет к переполнению регистров процессора и исключительным ситуациям. Поэтому в практической работе применяется логарифмическая арифметика (используются логарифмы вероятностей, а не их непосредственные значения).

Декодирование скрытой марковской модели. Процесс декодирования СММ позволяет определить: какова вероятность того, что наблюдаемая входная последовательность векторов $[(\{x)]_1, x_2, ..., x_T\}$ могла быть сгенерирована данной моделью, и соответствующую наиболее вероятную цепочку состояний. Для решения данной задачи применяется алгоритм максимума правдоподобия (Витерби). Последовательность действий на одном шаге декодирования изображена на рисунке 2.



$$p_{21}(t) = p_1(t-1) + a_{12}$$

$$p_{22}(t) = p_2(t-1) + a_{22}$$

$$p_{2n}(t) = p_n(t-1) + a_{n2}$$

$$p_2(t) = \max[p_{21}(t), p_{22}(t), \ldots, p_{2n}(t)]$$

Рис. 2. Алгоритм декодирования Витерби

В момент времени осуществляется переход в состояние i из *всех* предыдущих N состояний, после чего выбирается цепочка состояний, имеющая максимальную суммарную вероятность в моменты времени $t-1$. Время выполнения алгоритма пропорционально длине последовательности T и квадрату количества состояний N, алгоритм имеет сложность $O(N^2T)$.

Более подробное описание алгоритмов обучения можно найти в [2].

Анализ результатов. Для достижения поставленной цели – изучение свойств скрытых марковских моделей на примере решения задачи распознавания раздельной речи – был в полном объеме реализован аппарат СММ и проведено тестирование для обучающей выборки, составленной тремя дикторами (мужские голоса).

В качестве параметрического вектора использовались кепстральные коэффициенты (MFCC) и коэффициенты линейного предсказания (LPC) [3].

Для обучения и тестирования использовалась выборка из 10 слов (цифр) по 24 образца (8 для каждого диктора). Для каждого слова обучалась отдельная скрытая марковская модель (итого 10 моделей). Цель эксперимента – выяснить, при каких параметрах СММ ошибка обобщения будет минимальной. Другими словами, в процессе тестирования каждой модели необходимо *максимизировать* разность вероятностей декодирования «своих» и ложных образцов.

Исследовались следующие параметры:

– число состояний модели (N);

– число компонент гауссовой смеси (M);

– параметрический вектор (MFCC или LPC);

В таблицах 1–4 приведены наиболее характерные результаты тестирования системы, позволяющие сделать оценку зависимости точности распознавания от применяемого преобразования и количества состояний системы.

Таблица 1

Результаты тестирования для MFCC, N = 4, M = 16

Слово\СММ	Один	Два	Три	Четыре	Пять	Шесть	Семь	Восемь	Девять	Ноль
Один	**0,00**	–5,50	–7,70	–6,78	–8,41	–18,06	–12,52	–7,73	–5,78	–13,46
Два	–3,65	**0,00**	–8,26	–7,23	–8,58	–21,21	–14,44	–8,31	–5,58	–12,75
Три	–6,76	–7,65	**0,00**	–7,35	–7,73	–17,35	–15,41	–6,18	–7,06	–14,73
Четыре	–6,05	–8,29	–8,39	**0,00**	–8,77	–12,06	–7,57	–9,32	–7,15	–13,44
Пять	–6,05	–9,82	–8,17	–5,36	**0,00**	–8,45	–6,34	–4,46	–4,89	–13,33
Шесть	–10,43	–15,68	–12,81	–8,98	–7,30	**0,00**	–7,58	–9,67	–7,48	–14,52
Семь	–12,95	–14,84	–12,07	–8,00	–10,15	–11,06	**0,00**	–6,99	–8,82	–14,81
Восемь	–7,66	–10,01	–7,41	–7,99	–7,68	–12,61	–6,99	**0,00**	–7,44	–13,46
Девять	–7,30	–11,66	–9,29	–4,95	–7,10	–9,40	–5,41	–5,30	**0,00**	–13,80
Ноль	–13,47	–19,61	–14,46	–14,30	–13,22	–11,76	–11,86	–9,94	–14,51	**0,00**

Таблица 2

Результаты тестирования для MFCC, N = 8, M = 16

Слово\СММ	Один	Два	Три	Четыре	Пять	Шесть	Семь	Восемь	Девять	Ноль
Один	**0,00**	–6,11	–9,09	–7,56	–9,67	–18,36	–16,31	–12,12	–7,53	–16,30
Два	–5,43	**0,00**	–9,96	–8,63	–11,36	–22,09	–19,07	–11,97	–7,43	–15,60
Три	–8,07	–7,86	**0,00**	–7,96	–8,69	–16,92	–17,62	–7,66	–7,97	–17,05
Четыре	–6,75	–9,51	–9,62	**0,00**	–10,06	–11,81	–10,14	–10,38	–8,31	–18,09
Пять	–7,41	–10,48	–8,90	–7,54	**0,00**	–9,41	–7,68	–6,06	–7,59	–14,80
Шесть	–11,85	–14,43	–13,51	–9,76	–8,52	**0,00**	–8,12	–11,12	–10,37	–19,41
Семь	–13,53	–16,06	–12,91	–9,06	–10,75	–11,13	**0,00**	–7,75	–10,80	–18,94
Восемь	–8,13	–11,10	–7,89	–8,70	–7,98	–12,83	–7,74	**0,00**	–7,50	–14,70
Девять	–7,33	–12,34	–9,52	–5,16	–7,39	–8,78	–6,35	–5,41	**0,00**	–14,80
Ноль	–13,92	–17,44	–16,10	–14,16	–13,82	–11,69	–12,61	–9,82	–14,60	**0,00**

Таблица 3

Результаты тестирования для LPC, N = 4, M = 4

Слово\СММ	Один	Два	Три	Четыре	Пять	Шесть	Семь	Восемь	Девять	Ноль
Один	**0,00**	–9,41	–14,01	–12,93	–11,97	–23,96	–19,64	–14,83	–12,82	–25,70
Два	–7,76	**0,00**	–17,32	–17,58	–17,43	–28,92	–23,64	–20,59	–17,50	–26,15
Три	–9,78	–9,68	**0,00**	–9,10	–12,79	–23,93	–21,89	–14,20	–12,57	–22,63
Четыре	–14,00	–15,53	–17,64	**0,00**	–13,51	–20,69	–14,09	–17,38	–18,88	–36,03
Пять	–13,85	–14,18	–13,14	–8,62	**0,00**	–13,08	–10,27	–8,09	–9,09	–24,69
Шесть	–26,01	–27,07	–25,49	–20,09	–17,03	**0,00**	–16,84	–22,99	–24,31	–46,94
Семь	–20,80	–22,15	–20,05	–10,62	–14,45	–20,41	**0,00**	–12,25	–24,53	–46,15
Восемь	–12,24	–15,17	–10,98	–11,34	–9,09	–19,21	–11,62	**0,00**	–12,68	–21,22
Девять	–12,29	–15,66	–12,09	–6,00	–6,66	–13,11	–8,64	–7,31	**0,00**	–18,68
Ноль	–23,24	–22,51	–26,76	–21,27	–18,28	–19,83	–15,89	–16,80	–27,20	**0,00**

Таблица 4

Результаты тестирования для LPC, $N = 8$, $M = 4$

Слово\СММ	Один	Два	Три	Четыре	Пять	Шесть	Семь	Восемь	Девять	Ноль
Один	0,00	–12,75	–27,61	–16,01	–16,77	–42,71	–24,15	–21,32	–16,80	–42,87
Два	–16,04	0,00	–31,32	–21,94	–22,81	–48,06	–27,56	–25,08	–20,72	–43,05
Три	–14,67	–12,95	0,00	–11,52	–16,50	–40,89	–23,77	–18,98	–12,72	–34,50
Четыре	–21,94	–21,81	–32,09	0,00	–19,95	–34,73	–17,96	–20,84	–24,10	–66,94
Пять	–19,04	–20,26	–27,16	–14,30	0,00	–25,31	–19,31	–15,17	–17,76	–34,96
Шесть	–36,66	–37,37	–47,42	–22,36	–25,30	0,00	–25,53	–30,44	–29,24	–55,73
Семь	–27,62	–29,32	–43,77	–16,36	–22,47	–32,82	0,00	–16,01	–30,24	–75,26
Восемь	–16,60	–20,36	–22,52	–13,07	–14,53	–30,43	–14,66	0,00	–17,49	–35,84
Девять	–15,47	–21,08	–22,58	–11,05	–10,58	–21,79	–10,13	–9,96	0,00	–30,70
Ноль	–29,60	–30,94	–32,49	–26,09	–22,69	–27,34	–20,33	–20,21	–26,06	0,00

Каждый вертикальный столбец соответствует отдельной скрытой марковской модели. Горизонтальные строки соответствуют тестовым образцам. На пересечении i-той строки и j-того столбца находится усредненная логарифмическая вероятность декодирования j-той СММ i-того слова (чем меньше значением, тем меньше вероятность декодирования i-того образца j-той моделью, т.е. более высокая точность распознавания).

Проведенные тесты показали, что применение LPC-преобразования позволяет более высокие точности распознавания при меньшем количестве компонент гауссовой смеси (4 против 16), чем с помощью MFCC. Однако в некоторых случаях MFCC предпочтительней.

Наиболее оптимальное число состояний модели – 8 при средней длительности сигнала 0,5 сек. Дальнейшее увеличение количества состояний приводит к переобучению системы.

Итоговая достоверность достигла 95%.

Заключение. В ходе работы был реализован аппарат скрытых марковских моделей применительно к задаче распознавания раздельной речи. Получены практические результаты для параметрических векторов, основанных на кепстральных коэффициентах (MFCC) и коэффициентах линейного предсказания (LPC), и соответствующие оценки эффективности работы системы для различных параметров модели.

Дальнейшие исследования планируется проводить в нескольких направлениях:

1. Для повышения точности распознавания требуется применять более сложный параметрический вектор, например, объединить LPC, MFCC и ряд других преобразований. Однако увеличение размерности вектора приведет к значительным вычислительным затратам, в связи с чем предполагается использовать нелинейные преобразования (например, искусственные нейронные сети), позволяющие понизить размерность входного вектора.

2. Разработка и реализация алгоритмов преобразования обеспечат дикторо-независимость системы.

3. Возрастание размера обучающей выборки приведет к существенному увеличению времени обучения и тестирования, в результате может потребоваться применение современных технологий параллелизма, в частности, вычислений на графических процессорах (технология CUDA).

Библиографический список

1. Vaseghi, Saeed V. Advanced digital signal processing and noise reduction. – 3ed. – Chichester, 2006.

2. Hidden Markov Model Toolkit Book. – Cambridge University Engineering Department, 2001–2009.

3. Vaidyanathan P.P. The Theory of Linear Prediction. – California, 2008.