

## СКРЫТЫЕ МАРКОВСКИЕ МОДЕЛИ С ОДНОМЕРНОЙ ТОПОЛОГИЕЙ В ЗАДАЧЕ РАСПОЗНАВАНИЯ ЛИЦ

Т. А. ГУЛЬТЯЕВА<sup>♣</sup>, А. А. ПОПОВ<sup>♦</sup>

В данной статье представлен один из статистических методов распознавания лиц людей по их изображениям – скрытые марковские модели (СММ) с дискретным временем. На примере одномерной СММ рассмотрены различные методы извлечения наблюдений и выбора элементов СММ, влияющие на описательную способность модели.

### 1. ВВЕДЕНИЕ

Задача распознавания лиц актуальна как в области интеллектуальных сред (например, обеспечение доступа к каким – либо ресурсам, интерпретация поведения человека, естественный интерфейс человек-компьютер) так и в системах безопасности. Под распознаванием лиц будем понимать идентификацию изображения лица человека с какой-либо известной персоной из базы данных.

СММ успешно применяются для обработки аудио сигналов, речи, изображений, распознавания текста, в биомедицине и других областях [1].

### 2. ПОСТАНОВКА ЗАДАЧИ

Цифровое изображение – суть случайный двумерный дискретный сигнал, который наблюдается системой. Последовательность наблюдений  $O = o_1, o_2, \dots, o_T$ , где  $o_t$  – сигналы, упорядоченные по пространственным отношениям, может извлекаться из изображения различными способами. В силу этого описательные способности полученных моделей могут различаться.

Простейшее представление изображения лица – в виде последовательности наблюдаемых векторов, состоящих из значений пикселей, являющихся горизонтальными полосами размером  $X \times L_y$ , которые извлекаются сверху вниз с перекрытием в  $M_y$  пикселей (см. рис. 1). С целью уменьшения размерности пространства наблюдений и снижения вычислительной сложности используют некоторые преобразования (например, ДКП). Далее данный ме-

---

♣ Инженер-программист, магистр кафедры программных систем и баз данных

♦ Зав. кафедры программных систем и баз данных, профессор, д.-р техн. наук

тод представления изображения будем обозначать как SWB (Scan Window by Band – окно сканирования полосой).

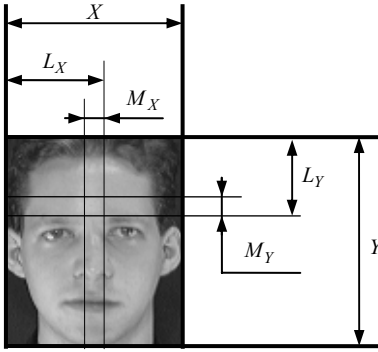


Рис. 1. Геометрические параметры изображения

$M_X$  пикселей по горизонтали и  $M_Y$  пикселей по вертикали (см. рис. 1). Число наблюдений определяется как  $T = T_X \cdot T_Y$ .

Элементами CMM являются:

1. Конечное множество скрытых состояний  $S = \{s_1, s_2, \dots, s_N\}$ .
2. Конечное множество наблюдаемых состояний (или дискретный алфавит)  $V = \{v_1, v_2, \dots, v_M\}$ .

3. Матрица переходных вероятностей  $A = \{a_{ij}\}$ , где  $a_{ij} = P(q_t = S_j | q_{t-1} = S_i)$ ,  $i, j = \overline{1, N}$ ,  $t = \overline{1, T}$  ассоциируема со стационарной марковской цепью на пространстве скрытых состояний. Здесь  $q_t$  – состояние модели в момент времени  $t$ . Пример ненаблюдаемой марковской цепь в случае 1D CMM представлен на рис. 2.

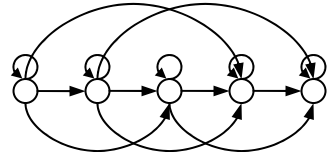


Рис. 2. Марковская цепь, соответствующая 1D CMM

4. Начальная вероятность состояний  $\pi = \{\pi_i\}$ ,  $\pi_i = P(q_1 = S_i)$ ,  $i = \overline{1, N}$ .

5. Матрица эмиссий  $B = \{b_i(k)\}$ ,  $b_i(k) = P(o_t = v_k | q_t = s_i)$ ,  $i = \overline{1, N}$ ,  $k = \overline{1, M}$ ,  $t = \overline{1, T}$ .

В задачах распознавания изображений используют либо дискретные модели, либо полунепрерывные СММ (непрерывное пространство наблюдаемых состояний, но дискретное пространство скрытых состояний). В первом случае получают более грубые модели, поскольку происходит дискретизация непрерывного сигнала (т. е.  $o_t$  является одним из  $u_k$ ), чем во втором, когда используется вероятностное описание наблюдаемых состояний. При этом вероятность нахождения некоторого наблюдения  $o_t$  в наблюдаемом состоянии  $m$  описывается нормальной функцией распределения  $N(o_t, \mu_m, \Sigma_m)$ , где параметры  $\mu_m$  и  $\Sigma_m$  распределения зависят от того, в каком скрытом состоянии находится  $o_t$ . Учитывая все вышесказанное, получаем:

$$b_i(t) = \sum_{m=1}^{M_i} c_{im} N(o_t, \mu_{im}, \Sigma_{im}),$$
 где  $c_{im}$  – коэффициент  $k$ -ой смеси  $i$ -ого состояния,  $M_i$  – число наблюдаемых состояний, описывающих  $i$ -ое ненаблюдаемое состояние.

Начиная с [3], где впервые было предложено использовать СММ для задачи распознавания, и в дальнейших работах на эту тему, наблюдаемые состояния  $V$  выбирались (в случае полунепрерывных СММ) как некоторые геометрические классы в пространстве наблюдений  $O$ . Далее будем обозначать данный метод как GB (General Base – общий базис).

Однако для наилучшего описания  $b_i(t)$  мы предлагаем множество  $V$  выбирать таким образом, чтобы только  $M_i$  конкретных наблюдаемых состояний описывали скрытое состояние  $s_i$ . Этот метод обозначим как PB (Private Base – частный базис). На рис. 3 приведены гистограммы распределения второго коэффициента ДКП наблюдений, попавших в первое состояние, и распределение  $b_1(t)$  при трех смесях (1D СММ).

На этапе обучения каждая модель  $\lambda_r = (\pi_r, A_r, B_r)$ ,  $r = 1, R$  ( $R$  – число персон, подлежащих распознаванию) настраивается на свои обучающие изображения в результате максимизации функций  $P(O^{(r)}, Q | \lambda_r)$  и  $P(O^{(r)} | \lambda_r)$ . Решение

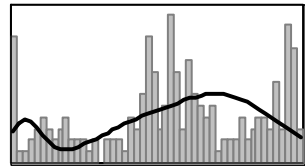


Рис. 3.1. GB Метод

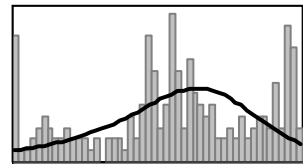


Рис. 3.2. PB Метод

первой подзадачи, т.е. определение максимизирующей последовательности скрытых состояний при фиксированной последовательности наблюдаемых состояний, дает алгоритм динамического программирования Витерби [3],[4]. Модификация EM алгоритма (алгоритм Баум–Уэлша) используется для решения второй подзадачи [3],[5]. В результате обучения полученная модель  $\lambda_r$  может быть использована в качестве генератора наблюдаемой последовательности  $O^{(r)}$  (извлеченных из обучающих изображений), а также при распознавании сигналов (тестовых изображений) в некотором смысле близких к  $O^{(r)}$ .

### 3. РЕЗУЛЬТАТЫ

Исследования проводились на базе изображений ORL<sup>1</sup>. Обучение каждой модели  $\lambda_r$  происходило по 5 случайно выбранным фотографиям данной персоны и их зеркальным отражениями (что позволяет увеличить в среднем процента верно распознанных изображений на 2%). Распознаванию подвергались не вошедшие в обучающий набор изображения. Результаты усреднялись по 10 выборкам.

На рис. 4 представлено изменение процента распознавания в зависимости от метода извлечения наблюдений, а также от способа описания  $b_i(t)$  при различном числе смесей  $M_i$ ,  $M = N \cdot M_i$ .

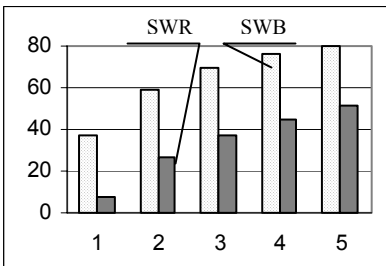


Рис. 4.1. Процент распознавания, GB метод

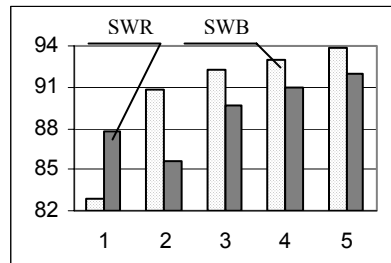


Рис. 4.2. Процент распознавания, PB метод

На рис. 5 и 6 для метода SWB приведены характеры зависимости  $P^*(O^{(r)} | \lambda_z)$  от числа смесей, где  $O^{(r)}$  – наблюдения из обучающего либо

<sup>1</sup>Данная база состоит из 400 полутонковых изображений лиц различного ракурса 40 персон, <http://www.orl.co.uk/facedatabase.html>

тестового наборов.  $P^*(O^{(r)} | \lambda_z)$  – это минимальная вероятность генерации моделью  $\lambda_z$  наблюдений, принадлежащие данной персоне ( $z = r$ ) и не принадлежащих ( $z \neq r$ ) ей.

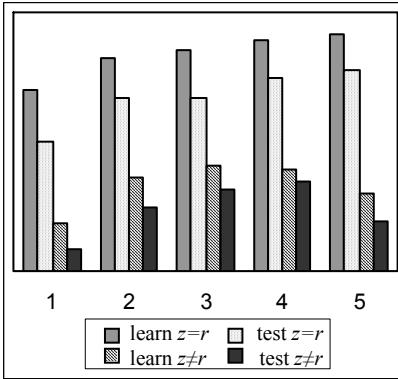


Рис. 5. Вероятность генерации наблюдений, GB метод

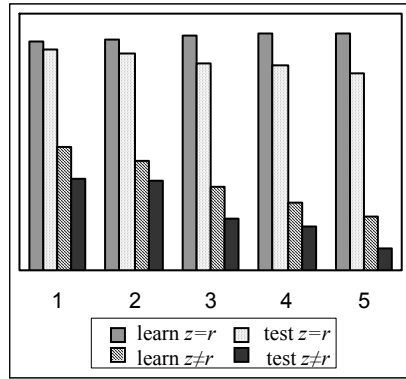


Рис. 6. Вероятность генерации наблюдений, PB метод

Сравнивая рис. 4.1 и рис. 5 можно сделать вывод, что при более точном описании моделью обучающих данных, вероятность сгенерировать тестовые изображения, принадлежащие данной персоне ( $z = r$ ) и изображения, не принадлежащие данной персоне ( $z \neq r$ ) увеличивается при увеличении числа смесей  $M_i$ . Следует обратить внимание на то, что начиная только с  $M_i = 4$  вероятность генерации наблюдений, являющихся ложными для модели, начинает снижаться. Как и предполагалось, модель с общим базисом (GB) недостаточно точно описывает тестовые изображения, что негативно сказывается на проценте верно распознанных изображений (см. рис. 4.1).

СММ с частным базисом (PB) (см. рис. 4.2 и рис. 6) настраивается более точно на воспроизведение обучающих наблюдений. Вероятность генерации ложных наблюдений с увеличением  $M_i$  снижается. При этом вероятность генерации своих тестовых наблюдений снижается, что говорит возможной переобученности модели.

Аналогичные результаты наблюдаются и для метода SWR.

При выборе количества смесей необходимо иметь ввиду, что вычислительные затраты линейно пропорциональны их числу (см. рис. 7).

Следует отметить, что все эксперименты, приведенные выше, проводились для примерно равного числа наблюдений ( $T \approx 110$ ). Таким образом, имеется возможность сравнить производительность систем, использующих методы SWB и SWR. По времени распознавания, SWR быстрее SWB в среднем в 1.2 раза, в то же время, процент распознавания ниже.

В других экспериментах было показано, что, увеличивая должным образом число наблюдений  $T$ , возможно повысить процент распознавания.

## ЗАКЛЮЧЕНИЕ

Использование метода PB позволяет повысить процент распознавания. При этом уменьшается и время, необходимое на распознавание одного изображения по сравнению с классическим методом GB (см. рис. 7).

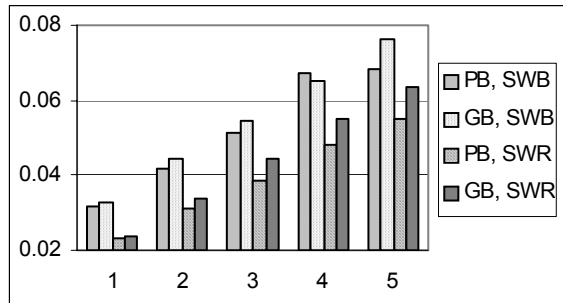


Рис. 7. Зависимость времени распознавания (с) одного изображения от числа смесей

[1] **Cappe O.** *Ten years of HMM*// CNRS, LTCI & ENST, Dpt. TSI <http://www.tsi.enst.fr/~cappe/>

[2] **Samaria F. S.** *Face Recognition Using Hidden Markov Models*, PhD Thesis, University of Cambridge, 1994.

[3] **Rabiner L. R.** *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*// Proceedings of the IEEE, 1989, vol. 77, no. 2, p. 257-285.

[4] **Kundu A., Bayya A.** *Speech Recognition Using Hybrid Hidden Markov Model and NN Classifier*// International journal of speech technology, 1998, vol. 2, p. 227-240.

[5] **Bilmes J. A.** *A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov models*, Technical Report 97-021, International Computer Science Institute Berkeley CA, 1998.