Скрытые марковские модели

Семинар был посвящен обсуждению нового для нас аппарата скрытых марковских моделей.

Скрытая марковская модель (СММ) — статистическая модель, имитирующая работу процесса, похожего на марковский процесс с неизвестными параметрами, и задачей ставится разгадывание неизвестных параметров на основе наблюдаемых. Полученные параметры могут быть использованы в дальнейшем анализе, например, для распознавания образов.

В скрытой марковской модели мы можем следить лишь за переменными, на которые оказывает влияние данное *скрытое* состояние цепи Маркова. Каждому состоянию соответствует некоторое вероятностное распределение на множестве всех возможных выходных значений.

Определим параметры, описывающие поведение СММ.

- T — длина последовательности наблюдений,
- N — количество скрытых состояний цепи,
- M — количество наблюдаемых состояний,
- $Q = \{q_1, q_2, \ldots, q_T\}$ - последовательность скрытых состояний цепи (q_t — скрытое состояние цепи в момент времени n),
- $V = \{v_1, v_2, \ldots, v_M\}$ — множество наблюдаемых состояний,
- $A = \{a_{ij}\}, a_{ij} = \mathbf{P}(q_{t+1} = S_j \mid q_t = S_i)$ — матрица переходных вероятностей скрытой ЦМ,
- $B = \{b_j(k)\}, b_j(k) = \mathbf{P}(v_k \text{ в момент } t \mid q_t = S_i)$ — вероятность наблюдения (эмиссии) из состояния j,
- $\pi = \{\pi_i\}, \pi_i = \mathbf{P}(q_1 = S_i)$ — начальное распределение цепи.

Последовательность наблюдений обозначается через $O = \{O_1, \ldots, O_T\}$. Введем также обозначение $\lambda = (A, B, \pi)$.

Для СММ можно сформулировать 3 вопроса, комбинируя которые можно решать различные прикладные задачи.

Задача 1. По известным наблюдениям $O = \{O_1, \ldots, O_T\}$ и параметрам $\lambda = (A, B, \pi)$ найти вероятность $\mathbf{P}(O \mid \lambda)$.

Задача 2. По известным наблюдениям $O = \{O_1, \ldots, O_T\}$ и параметрам $\lambda = (A, B, \pi)$ найти наиболее оптимальную (в некотором смысле) последовательность скрытых состояний $Q = \{q_1, q_2, \ldots, q_N\}$, стоящую за данными наблюдениями.

Задача 3 По известным наблюдениям $O = \{O_1, \ldots, O_T\}$ подобрать параметры $\lambda = (A, B, \pi)$, максимизирующие $\mathbf{P}(O \mid \lambda)$.

Проиллюстрируем, как эти задачи могут применяться в распознавании речи.

Пример. Пусть W — словарь слов, каждое из которых мы должны научиться распознавать из звукового сигнала. Для этого для каждого слова $w \in W$ мы создаем СММ с N скрытыми состояниями. Чтобы обучить каждую из них, в качестве наблюдаемой последовательности нужно использовать показания, принимающие одно из M значений. Чтобы преобразовать исходный звуковой сигнал в такую дискретную последовательность, проделывают процедуру *квантования* сигнала. В результате процедуры получается выборка длины T, каждый элемент которой принимает одно из M значений.

Имея теперь выборку наблюдений, можно решить задачу 3 и для каждой СММ подобрать набор параметров λ_w.

Далее, решая задачу 2 для построенных СММ, можно придавать смысл скрытым состояниям и более тонко настраивать модель: например, менять количество скрытых состояний.

Наконец, имея обученные СММ для каждого слова из словаря, для нового поступившего сигнала можно находить вероятности $\mathbf{P}(O \mid \lambda_w)$ для всех слов w из словаря (это задача 1), и считать моделью для сигнала то слово, для которого эта вероятность максимальна.

На семинаре мы рассмотрели алгоритмы решения первых двух задач.

Решение задачи 1. Алгоритм прямого-обратного хода

Введем переменную прямого хода

$$\alpha_t(i) = P\left(O_1 O_2 \cdots O_t, q_t = S_i \mid \lambda\right).$$

Ее значения для всех $t = 1, \ldots, T$ и $i = 1, \ldots, N$ можно найти итеративно:

1. Инициализация: $\alpha_1(i) = \pi_i b_i\left(O_1\right), \quad 1 \leq i \leq N;$
2. Индукция: $\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] b_j\left(O_{t+1}\right), \quad 1 \leq t \leq T-1$ и $1 \leq j \leq N$.

Искомую вероятность теперь можно найти как

$$\mathbf{P}(O \mid \lambda) = \sum_{i=1}^{N} \alpha_T(i).$$

Трудоемкость алгоритма составляет $O(N^2 T)$.

Решение задачи 2. Алгоритм Витерби (Viterbi)

Чтобы найти наилучшую последовательность скрытых состояний

$$Q = \{q_1, q_2, \ldots, q_N\},$$

которая объясняет наблюдения $O = \{O_1, \ldots, O_T\}$, определим величину

$$\delta_t(i) = \max_{q_1, q_2, \ldots, q_{t-1}} \mathbf{P}\left(q_1 q_2 \ldots q_t = S_i, O_1 O_2 \ldots O_t \mid \lambda\right).$$

Она имеет смысл максимальной вероятности по всем путям, которые заканчиваются в момент времени t в скрытом состоянии S_i. Чтобы восстановить сам путь, будем использовать переменные $\psi_t(i)$.

1. Инициализация:

$$\begin{aligned} \delta_1(i) &= \pi_i b_i\left(O_1\right), \quad 1 \le i \le N \\ \psi_1(i) &= 0. \end{aligned}$$

2. Индукция:

$$\begin{aligned} \delta_t(j) &= \max_{1 \le i \le N} \left[\delta_{t-1}(i) a_{ij}\right] b_j\left(O_t\right), \quad 2 \le t \le T, 1 \le j \le N \\ \psi_t(j) &= \operatorname*{argmax}_{1 \le i \le N} \left[\delta_{t-1}(i) a_{ij}\right], \quad 2 \le t \le T, 1 \le j \le N. \end{aligned} \tag{1}$$

Теперь

$$\begin{aligned} \mathbf{P}^* &= \max_{1 \le i \le N} \left[\delta_T(i)\right], \\ q_T^* &= \operatorname*{argmax}_{1 \le i \le N} \left[\delta_T(i)\right]. \end{aligned}$$

Восстановим исходный путь:

$$q_t^* = \psi_{t+1}\left(q_{t+1}^*\right), \quad t = T-1, T-2, \ldots, 1.$$

Используя аналогичный подход (динамического программирования), можно решить и задачу 3. Этот алгоритм называется алгоритмом Баума — Уэлша (Baum — Welch).