Versteckte Markov-Modelle

Wolfgang Mulzer

1 Modell

Ein *verstecktes Markov-Modell* (Hidden Markov Model, HMM) hat folgende Bestandteile:

1. Eine endliche Zustandsmenge Q,
2. ein endliches Ausgabealphabet Σ,
3. eine Anfangsverteilung a_q für $q \in Q$ mit $a_q \in [0,1]$ und $\sum_{q \in Q} a_q = 1$,
4. Übergangsverteilungen t_{qr} für $q, r \in Q$ mit $t_{qr} \in [0,1]$ und $\sum_{r \in Q} t_{qr} = 1$ für jeden Zustand $q \in Q$, und
5. Ausgabeverteilungen $o_{q\sigma}$ für $q \in Q$ und $\sigma \in \Sigma$ mit $o_{q\sigma} \in [0,1]$ und $\sum_{\sigma \in \Sigma} o_{q\sigma} = 1$ für jeden Zustand $q \in Q$.

Ein verstecktes Markov-Modell modelliert den folgenden Zufallsprozess.

1. Wähle einen zufälligen Zustand q_0 gemäß der Anfangsverteilung a_q.
2. Sei q_i der aktuelle Zustand. Wähle eine zufällige Ausgabe σ_i gemäß der Ausgabeverteilung $o_{q_i\sigma}$.
3. Wähle einen zufälligen Nachfolgezustand r gemäß der Übergangsverteilung t_{q_ir}. Setze $q_{i+1} = r$ und wiederhole.

2 Der Viterbi-Algorithmus

Nun stellt sich folgendes Problem. Angenommen, wir beobachten eine zufällige Ausgabe $\sigma_0, \sigma_1, \dots, \sigma_{\ell-1}$. Welche Zustandsfolge hat diese Ausgabe erzeugt?

Natürlich ist diese Frage so nicht zu beantworten, da potentiell viele verschiedene Zustandsfolgen dieselbe Ausgabe produzieren können. Dennoch können wir eine sinnvolle Frage stellen. Die Ausgabe $\sigma_0, \sigma_1, \dots, \sigma_{\ell-1}$ induziert eine Wahrscheinlichkeitsverteilung auf den Zustandsfolgen der Länge ℓ, indem wir einer Folge $q_0, q_1, \dots, q_{\ell-1}$ die bedingte Wahrscheinlichkeit

$$\Pr[q_0, q_1, \dots, q_{\ell-1} \text{ Zustände } \mid \sigma_0, \sigma_1, \dots, \sigma_{\ell-1} \text{ Ausgabe}]$$

zuordnen. Nun suchen wir eine Zustandsfolge, die diese bedingte Wahrscheinlichkeit maximiert. Eine solche Zustandsfolge nennen wir auch eine *wahrscheinlichste Erklärung* für die Ausgabe $\sigma_0, \dots, \sigma_{\ell-1}$.

Mit dem Satz von Bayes sehen wir, dass eine Zustandsfolge die bedingte Wahrscheinlichkeit genau dann maximiert, wenn sie die Wahrscheinlichkeit

$$\Pr[\sigma_0, \ldots, \sigma_{\ell-1} \text{ Ausgabe } \mid q_0, \ldots, q_{\ell-1} \text{ Zustände}] \cdot \Pr[q_0, \ldots, q_{\ell-1} \text{ Zustände}]$$

maximiert. Diesen Ausdruck können wir schreiben als

$$o_{q_0\sigma_0} \cdot o_{q_1\sigma_1} \cdot \cdots \cdot o_{q_{\ell-1}\sigma_{\ell-1}} \cdot a_{q_0} \cdot t_{q_0 q_1} \cdot t_{q_1 q_2} \cdot \cdots \cdot t_{q_{\ell-2} q_{\ell-1}}.$$

Nun finden wir einen effizienten Algorithmus mit Hilfe von dynamischem Programmieren.

1. Finde geeignete Teilprobleme: Definiere $\mathtt{E}[q, i]$ als die maximale Wahrscheinlichkeit über alle Zustandsfolgen $q_0, q_1, \ldots, q_i = q$, dass das HMM die Zustandsfolge $q_0, q_1, \ldots, q_i = q$ ausführt und dabei die Ausgabe $\sigma_0, \sigma_1, \ldots, \sigma_i$ produziert.

2. Finde eine Rekursion für $\mathtt{E}[q, i]$: Zu Beginn ist die Wahrscheinlichkeit, dass wir uns im Zustand q befinden, a_q und die Wahrscheinlichkeit, dass σ_0 ausgegeben wird, ist $o_{q\sigma_0}$.

$$\mathtt{E}[q, 0] = a_q \cdot o_{q\sigma_0}, \qquad \text{für alle } q \in Q.$$

   Für die Rekursion unterscheiden wir, welches der vorletzte Zustand einer optimalen Zustandsfolge war, und wir nehmen das Maximum.

$$\mathtt{E}[q, i] = \max_{r \in Q} \{\mathtt{E}[r, i-1] \cdot t_{rq} \cdot o_{q\sigma_i}\}, \qquad \text{für } q \in Q \text{ und } i = 1, \ldots, \ell - 1.$$

3. Implementiere die Rekursion: Übung.

4. Finde eine optimale Zustandsfolge: Übung.

Man erhält eine Laufzeit von $O(\ell |Q|^2)$ und Platzbedarf $O(\ell |Q|)$.