Hidden Markov Models

Bartek Wilczyński

April 10th, 2018

Materiały do dzisiejszego wykładu

- Do dzisiejszego wykładu: `http://www.mimuw.edu.pl/~dojer/wobm/hmm.pdf`

Markov Models

- The model consists of a state space $Q \neq \emptyset$ (for our purposes Q is finite)
- and a transition probability matrix p_{ij} where $i, j \in Q$
- The model has no memory, the probability of moving from state i to j depends only on the state i.
- multiplying the matrix P, we can compute the change of the probability distribution as the model "steps" forward
- We are usually interested in stationary distributions π, such that $\pi \cdot P = P$

Hidden Markov Model

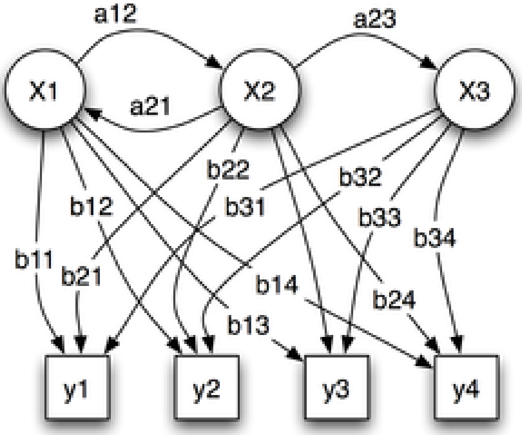

image (c) wikipedia

Hidden Markov Model - trajectory

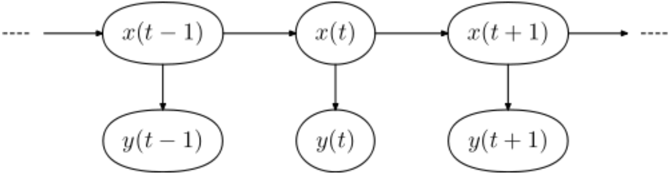

image (c) wikipedia

Hidden Markov Model - example

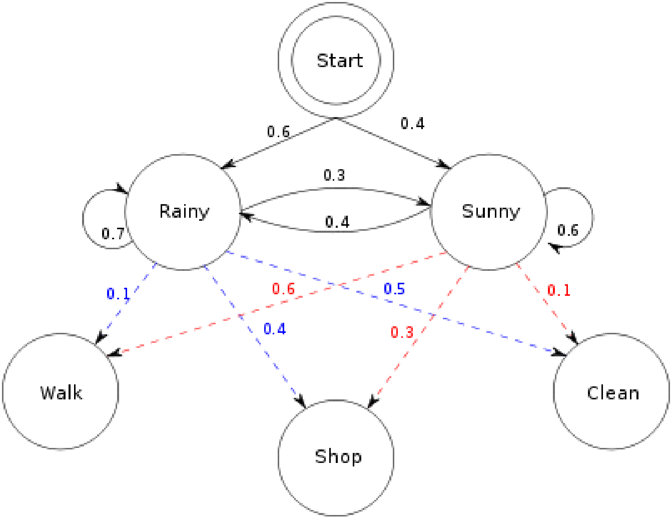

image (c) wikipedia

Reconstructing HMM trajectories

For any trajectory π, we can calculate the probability of emiting S

$$P(S, \pi) = \prod_{t=0}^{n-1} e_{\pi(t+1)}(S(t+1)) \cdot p_{\pi(t),\pi(t+1)},$$

Can we find the optimal trajectory π, given S?

$$P(S, \pi_*) = max\{P(S, \pi) \mid \pi \in Q^*, |\pi| = |S|\}.$$

Viterbi algorithm

We can use dynamic programming, filling in the $v(i,k)$ matrix

$$v(i,k) = max\{P(S[1..i],\pi) \mid \pi \in Q^i, \quad \pi(i) = k\}.$$

with the initial condition:

$$v(0,k) = \begin{cases} 1 & \text{gdy } k = k_0, \\ 0 & \text{gdy } k \neq k_0. \end{cases}$$

and step function:

$$v(i,k) = e_k(S(i)) \cdot \max_{l \in Q}[v(i-1,l) \cdot p_{l,k}].$$

To finally read out the seeked probability:

$$P(S,\pi_*) = \max_{k \in Q}[v(|S|,k)].$$

Estimating emission probabilities

Now, we can calculate the probability of emitting S, over all possible trajectories, with the Forward-method. The initial step is as follows:

$$f(0,k) = \begin{cases} 1 & \text{gdy } k = k_0, \\ 0 & \text{gdy } k \neq k_0. \end{cases}$$

Then, we make similar steps:

$$f(i,k) = e_k(S(i)) \cdot \sum_{l \in Q} f(i-1,l) \cdot p_{l,k}.$$

and finally we can calculate the total probability at the end:

$$P(S) = \sum_{k \in Q} f(|S|,k).$$

The same works backwards:

$$b(i,k) = \sum_{l \in Q} p_{k,l} \cdot e_l(S(i+1)) \cdot b(i+1,l).$$

Estimating emission probabilities

Putting it together, probability of being in state k at step i, given S:

$$P(\pi(i) = k \mid S) = \frac{P(\pi(i) = k \,\&\, S)}{P(S)} = \frac{f(i,k) \cdot b(i,k)}{P(S)}.$$

emission probabilities

Estimate of the Emission matrix:

$$e_k(x) = \frac{E_k(x)}{\sum_{y \in \Sigma} E_k(y)}.$$

Can be calculated using f and b

$$E_k(x) = \sum_{j=1}^{n} \sum_{i \in I_j(x)} \frac{f_{\mathcal{M}}^{(j)}(i,k) \cdot b_{\mathcal{M}}^{(j)}(i,k)}{P_{\mathcal{M}}(S_j)},$$

Calculating probabilities

Similarly the transition matrix:

$$p_{k,l} = \frac{P_{k,l}}{\sum_{q \in Q} P_{k,q}},$$

depends on f and b

$$P_{k,l} = \sum_{j=1}^{n} \sum_{i=1}^{|S_j|} \frac{f_{\mathcal{M}}^{(j)}(i,k) \cdot p_{k,l}^{\mathcal{M}} \cdot e_l^{\mathcal{M}}(S_j(i+1)) \cdot b_{\mathcal{M}}^{(j)}(i+1,l)}{P_{\mathcal{M}}(S_j)}.$$

Baum-Welch algorithm

- Suppose, we only know the word S and the sets Q and Σ. Can we estimate both p_{ij} and e_{ij}?
- We can start with random p_{ij}, e_{ij} and iteratively proceed as follows:
  - Calculate the estimates of being in each of states at each step using f, b and current estimates of e, p.
  - Find the optimal e, p, given current e, p, f, b
- This is an example of a known procedure called Expectation-Maximization
- It converges to a local optimum of the likelihood, because at every iteration, the likelihood cannot be decreased.

Extensions and applications

- We will discuss it in more depth next week, but the HMM model is very useful in describing sequence alignments and so-called sequence profiles
- It is relatively easy to extend this model for arbitrary emissions (e.g. Gaussian or multinomial), not necessarily from a discrete space of symbols. This is frequently used for modelling functional genomics data
- It is also quite a good model for segmentation of chromosomes based on different measurements along the genome