HMM for NLP

Probabilistic Graphical Models

Jerónimo Hernández-González

Template models (rev.)

What is a template model?

- $\boldsymbol{X}$ takes different values at each (discrete) time step
  $\boldsymbol{X}(t)$ is the random variable at time t
- Markov assumption

$$\boldsymbol{X}(t+1) \perp\!\!\!\perp \boldsymbol{X}(0), \ldots, \boldsymbol{X}(t-1) \mid \boldsymbol{X}(t)$$

- Stationary assumption (Time invariance or homogenous)

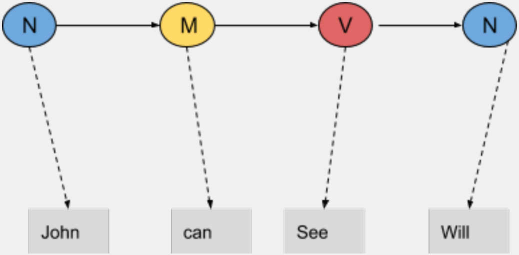
$$P(\boldsymbol{X}(t+1) \mid \boldsymbol{X}(t)), \text{ the same for all } t$$

- Use **conditional** Bayesian network to define $P(\boldsymbol{X}(t+1)|\boldsymbol{X}(t))$
  2-time slice Bayesian network, Dynamic Bayesian network, **Hidden Markov models**

Hidden Markov Model

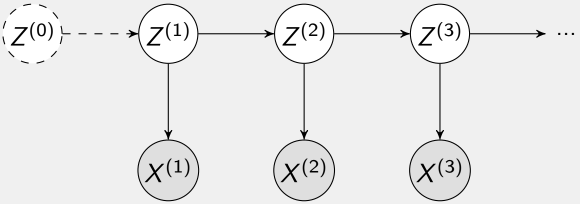
HMM

HMMs for NLP (e.g., POS-tagging)



Hidden Markov Model

HMM



$$p(\boldsymbol{Z}, \boldsymbol{X}) = p(Z^{(0)}) \cdot \prod_t p(Z^{(t)}|Z^{(t-1)}) \cdot \prod_t p(X^{(t)}|Z^{(t)})$$

Hidden Markov Model

HMM

$$p(\boldsymbol{Z}, \boldsymbol{X}) = p(Z^{(0)}) \cdot \prod_t p(Z^{(t)} | Z^{(t-1)}) \cdot \prod_t p(X^{(t)} | Z^{(t)})$$

We really want to know:

$$p(Z^{(t)} | \boldsymbol{x}^{(1:T)})$$

Forwards-Backwards algorithm

$$\gamma^{(t)}(j) = p(Z^{(t)} = j | \boldsymbol{x}^{(1:T)})$$

$$\gamma^{(t)}(j) \propto p(Z^{(t)} = j | \boldsymbol{x}^{(1:t)}) \cdot p(\boldsymbol{x}^{(t+1:T)} | Z^{(t)} = j)$$

$$\gamma^{(t)}(j) \propto \alpha^{(t)}(j) \cdot \beta^{(t)}(j)$$

Forwards-Backwards algorithm

$$\gamma^{(t)}(j) \propto \alpha^{(t)}(j) \cdot \beta^{(t)}(j)$$

Forwards pass:

$$\begin{aligned}\alpha^{(t)}(j) &= p(Z^{(t)} = j | \boldsymbol{x}^{(1:t)}) \\ &= p(Z^{(t)} = j | x^{(t)}, \boldsymbol{x}^{(1:t-1)}) \\ &\propto p(x^{(t)} | Z^{(t)} = j) \cdot p(Z^{(t)} = j | \boldsymbol{x}^{(1:t-1)})\end{aligned}$$

$$\begin{aligned}\alpha^{(t)}(j) \propto p(x^{(t)} | Z^{(t)} = j) \cdot \textstyle\sum_k p(Z^{(t)} = j | Z^{(t-1)} = k) \cdot \\ \cdot p(Z^{(t-1)} = k | \boldsymbol{x}^{(1:t-1)})\end{aligned}$$

$$\alpha^{(t)}(j) \propto p(x^{(t)} | Z^{(t)} = j) \cdot \sum_k p(Z^{(t)} = j | Z^{(t-1)} = k) \cdot \alpha^{(t-1)}(k)$$

» To compute $\alpha^{(t)}$, you need to compute first $\alpha^{(t-1)}$

Hidden Markov Model

HMM

$$p(\boldsymbol{Z}, \boldsymbol{X}) = p(Z^{(0)}) \cdot \prod_t p(Z^{(t)}|Z^{(t-1)}) \cdot \prod_t p(X^{(t)}|Z^{(t)})$$

We really want to know:

$$p(Z^{(t)}|\boldsymbol{x}^{(1:T)})$$

Forwards-Backwards algorithm

$$\gamma^{(t)}(j) = p(Z^{(t)} = j|\boldsymbol{x}^{(1:T)})$$

$$\gamma^{(t)}(j) \propto p(Z^{(t)} = j|\boldsymbol{x}^{(1:t)}) \cdot p(\boldsymbol{x}^{(t+1:T)}|Z^{(t)} = j)$$

$$\gamma^{(t)}(j) \propto \alpha^{(t)}(j) \cdot \beta^{(t)}(j)$$

Hidden Markov Model

Forwards-Backwards algorithm

$$\gamma^{(t)}(j) \propto \alpha^{(t)}(j) \cdot \beta^{(t)}(j)$$

Backwards pass:

$$\begin{aligned}\beta^{(t-1)}(j) &= p(\boldsymbol{x}^{(t:T)}|Z^{(t-1)} = j) \\ &= \textstyle\sum_k p(Z^{(t)} = k, x^{(t)}, \boldsymbol{x}^{(t+1:T)}|Z^{(t-1)} = j)\end{aligned}$$

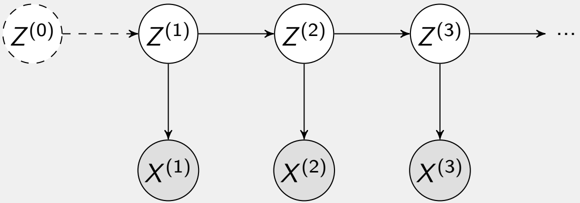
$$\beta^{(t-1)}(j) = \textstyle\sum_k p(\boldsymbol{x}^{(t+1:T)}|Z^{(t)} = k) \cdot p(x^{(t)}|Z^{(t)} = k) \cdot \\ \cdot p(Z^{(t)} = k|Z^{(t-1)} = j)$$

$$\beta^{(t-1)}(j) = \sum_k \beta^{(t)}(k) \cdot p(x^{(t)}|Z^{(t)} = k) \cdot p(Z^{(t)} = k|Z^{(t-1)} = j)$$

» To compute $\beta^{(t-1)}$, you need to compute first $\beta^{(t)}$

Hidden Markov Model

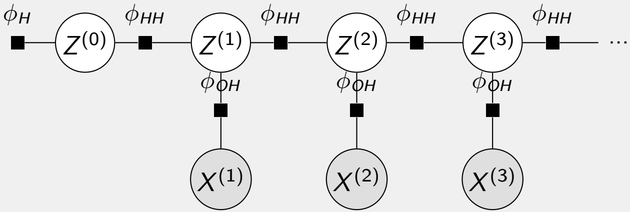
HMM



$$p(\boldsymbol{Z}, \boldsymbol{X}) = p(Z^{(0)}) \cdot \prod_t p(Z^{(t)} | Z^{(t-1)}) \cdot \prod_t p(X^{(t)} | Z^{(t)})$$

Hidden Markov Model

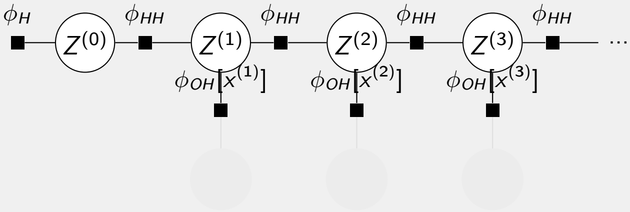
HMM



$$p(\boldsymbol{Z}, \boldsymbol{X}) = \phi_H(Z^{(0)}) \cdot \prod_t \phi_{HH}(Z^{(t)}, Z^{(t-1)}) \cdot \prod_t \phi_{OH}(X^{(t)}, Z^{(t)})$$

Hidden Markov Model

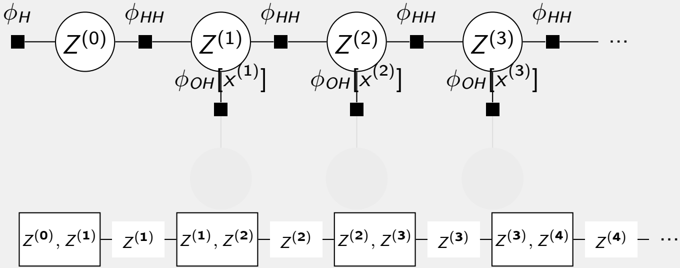
HMM



$$p(\boldsymbol{Z}|\boldsymbol{x}) \propto \phi_H(Z^{(0)}) \cdot \prod_t \phi_{HH}(Z^{(t)}, Z^{(t-1)}) \cdot \prod_t \phi_{OH}(X^{(t)} = x^{(t)}, Z^{(t)})$$

Hidden Markov Model

HMM



Hidden Markov Model

HMM

$$\boxed{Z^{(0)}, Z^{(1)}} - Z^{(1)} - \boxed{Z^{(1)}, Z^{(2)}} - Z^{(2)} - \boxed{Z^{(2)}, Z^{(3)}} - Z^{(3)} - \boxed{Z^{(3)}, Z^{(4)}} - Z^{(4)} - \cdots$$

$$\delta_{(t-1)\to t}(Z^{(t)}) = \textstyle\sum_{k\in\Omega_{Z^{(t-1)}}} \phi_{HH}(Z^{(t-1)} = k, Z^{(t)})\cdot$$
$$\cdot\phi_{OH}[x^{(t)}](Z^{(t)}) \cdot \delta_{(t-2)\to(t-1)}(Z^{(t-1)} = k)$$

$$\delta_{t\to(t-1)}(Z^{(t)}) = \textstyle\sum_{k\in\Omega_{Z^{(t+1)}}} \phi_{HH}(Z^{(t)}, Z^{(t+1)} = k)\cdot$$
$$\cdot\phi_{OH}[x^{(t+1)}](Z^{(t+1)} = k) \cdot \delta_{(t+1)\to t}(Z^{(t+1)} = k)$$

Hidden Markov Model

HMM

$$p(\boldsymbol{Z}, \boldsymbol{X}) = p(Z^{(0)}) \cdot \prod_t p(Z^{(t)}|Z^{(t-1)}) \cdot \prod_t p(X^{(t)}|Z^{(t)})$$

We really want to know:

$$p(Z^{(t)}|\boldsymbol{x}^{(1:T)})$$

Forwards-Backwards algorithm

$$\gamma^{(t)}(j) = p(Z^{(t)} = j|\boldsymbol{x}^{(1:T)})$$

$$\gamma^{(t)}(j) \propto p(Z^{(t)} = j|\boldsymbol{x}^{(1:t)}) \cdot p(\boldsymbol{x}^{(t+1:T)}|Z^{(t)} = j)$$

$$\gamma^{(t)}(j) \propto \alpha^{(t)}(j) \cdot \beta^{(t)}(j)$$

Hidden Markov Model

HMM

$$\boxed{Z^{(0)}, Z^{(1)}} - Z^{(1)} - \boxed{Z^{(1)}, Z^{(2)}} - Z^{(2)} - \boxed{Z^{(2)}, Z^{(3)}} - Z^{(3)} - \boxed{Z^{(3)}, Z^{(4)}} - Z^{(4)} - \cdots$$

Given the cluster factor:

$$\beta^{(t)}(Z^{(t)}, Z^{(t+1)}) = \phi_{HH}(Z^{(t)}, Z^{(t+1)}) \cdot \phi_{OH}[x^{(t+1)}](Z^{(t+1)}) \cdot \\ \cdot \delta_{(t-1)\to t}(Z^{(t)}) \cdot \delta_{(t+1)\to t}(Z^{(t+1)})$$

The marginal $p(Z^{(t)}|\boldsymbol{x}^{(1:T)})$ is obtained as follows:

$$\begin{aligned} p(Z^{(t)}|\boldsymbol{x}^{(1:T)}) &= \textstyle\sum_{k\in\Omega_{Z^{(t+1)}}} \beta^{(t)}(Z^{(t)}, Z^{(t+1)}) \\ &= \textstyle\sum_{k\in\Omega_{Z^{(t+1)}}} \phi_{HH}(Z^{(t)}, Z^{(t+1)}) \cdot \phi_{OH}[x^{(t+1)}](Z^{(t+1)}) \cdot \\ &\qquad \cdot \delta_{(t-1)\to t}(Z^{(t)}) \cdot \delta_{(t+1)\to t}(Z^{(t+1)}) \\ &= \delta_{(t-1)\to t}(Z^{(t)}) \cdot \delta_{t\to(t-1)}(Z^{(t)}) \end{aligned}$$

References

K. Murphy (2012) Chapter 17.4.3: The forwards-backwards algorithm. In *Machine Learning: a Probabilistic Perspective*.

HMM for NLP

Probabilistic Graphical Models

Jerónimo Hernández-González