Hidden Markov Models

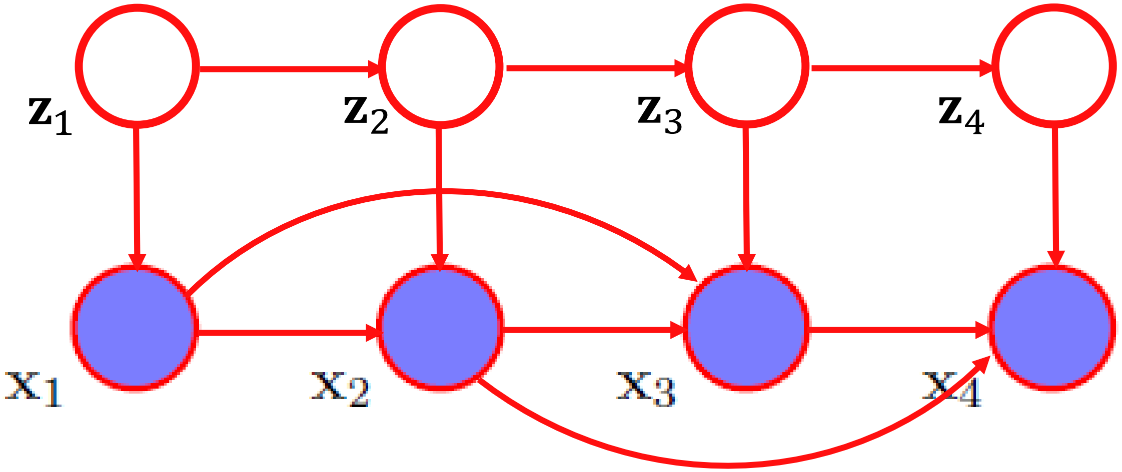
2019.01.25

浙江大學
Zhejiang University

Contents

Introduction

For large order Markov Chain, it's hard to evaluate the exponentially growing parameters.



Hidden Markov Model(HMM)

. . . . . .

$$\mathbf{z}_{n+1} \perp\!\!\!\perp \mathbf{z}_{n-1} \mid \mathbf{z}_n$$

$$p(\mathbf{x}) = p(\mathbf{x}_1)\prod_{n=2}^{N} p(\mathbf{x}_n|\mathbf{x}_{n-1})$$ First-order Markov Chain

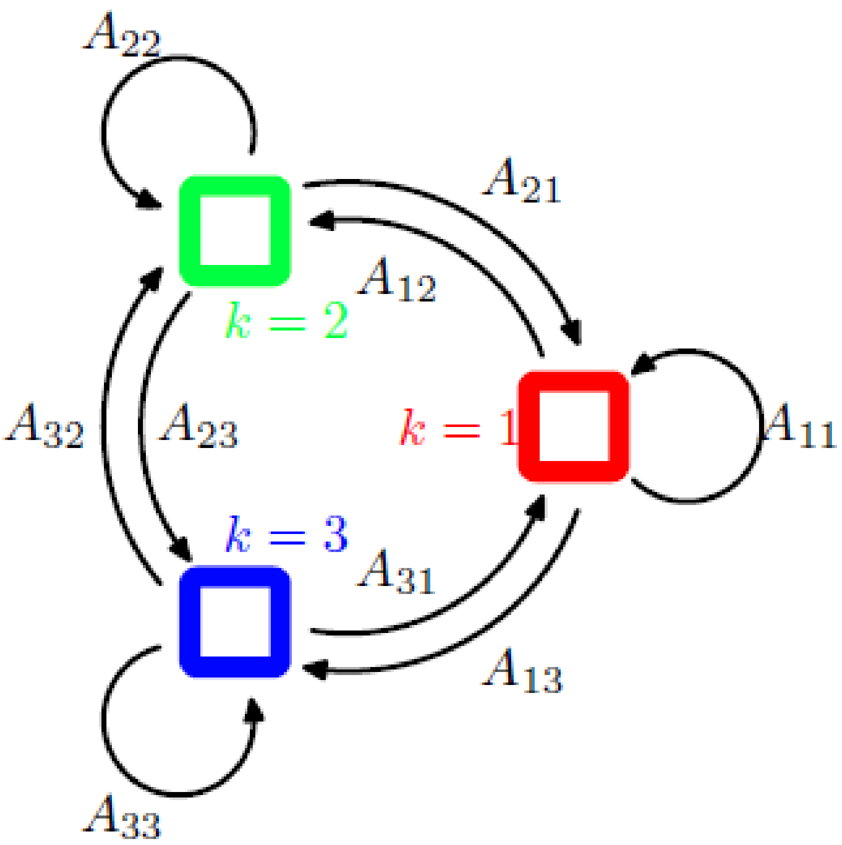
$$p(\mathbf{x}) = p(\mathbf{x}_1)p(\mathbf{x}_2|\mathbf{x}_1)\prod_{n=3}^{N} p(\mathbf{x}_n|\mathbf{x}_{n-1},\mathbf{x}_{n-2})$$ Second-order Markov Chain

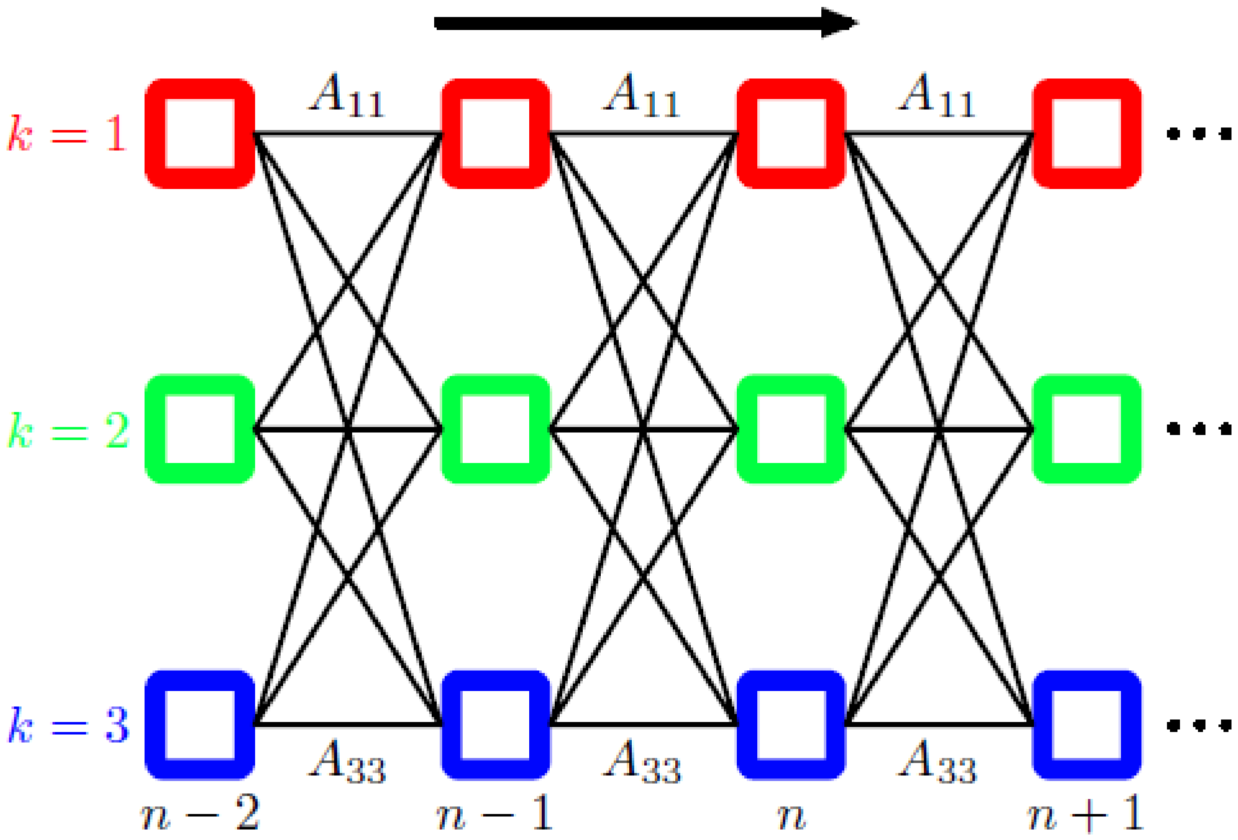
$$p(\mathbf{x},\mathbf{z}) = p(\mathbf{z}_1)\prod_{n=2}^{N} p(\mathbf{z}_n|\mathbf{z}_{n-1})\prod_{n=2}^{N} p(\mathbf{x}_n|\mathbf{z}_n)$$

transition

emission

Introduction – Transition Probability





Contents

浙江大學
Zhejiang University

Algorithms

1. Backward-Forward (Sum-product)
2. Maximum Likelihood ---- EM
3. Viterbi Algorithm

浙江大學
Zhejiang University

Algorithms – EM framework

Likelihood Function:

$$p(\mathbf{X}|\boldsymbol{\theta}) = \sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}).$$

In E step, we estimate $p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{old})$. Considering the relationship between $\mathbf{z}_n$ and $\mathbf{z}_{n-1}$, we define:

$$\gamma(z_{nk}) = \mathbb{E}[z_{nk}] = \sum_{\mathbf{z}} \gamma(\mathbf{z}) z_{nk}$$

$$\xi(z_{n-1,j}, z_{nk}) = \mathbb{E}[z_{n-1,j} z_{nk}] = \sum_{\mathbf{z}} \gamma(\mathbf{z}) z_{n-1,j} z_{nk}$$

In M step, we maximize:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$$

Here $\boldsymbol{\theta}$ represents parameters.

浙江大学 Zhejiang University

Algorithms – Forward-Backward

To seek an efficient procedure for evaluating the quantities $\gamma(z_{nk})$ and $\xi(z_{n-1,j}, z_{nk})$

$$
\begin{aligned}
p(\mathbf{X}|\mathbf{z}_n) &= p(\mathbf{x}_1,\ldots,\mathbf{x}_n|\mathbf{z}_n) \\
&\quad p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n) \\
p(\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}|\mathbf{x}_n,\mathbf{z}_n) &= p(\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}|\mathbf{z}_n) \\
p(\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}|\mathbf{z}_{n-1},\mathbf{z}_n) &= p(\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}|\mathbf{z}_{n-1}) \\
p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n,\mathbf{z}_{n+1}) &= p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_{n+1}) \\
p(\mathbf{x}_{n+2},\ldots,\mathbf{x}_N|\mathbf{z}_{n+1},\mathbf{x}_{n+1}) &= p(\mathbf{x}_{n+2},\ldots,\mathbf{x}_N|\mathbf{z}_{n+1}) \\
p(\mathbf{X}|\mathbf{z}_{n-1},\mathbf{z}_n) &= p(\mathbf{x}_1,\ldots,\mathbf{x}_{n-1}|\mathbf{z}_{n-1}) \\
&\quad p(\mathbf{x}_n|\mathbf{z}_n)p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n) \\
p(\mathbf{x}_{N+1}|\mathbf{X},\mathbf{z}_{N+1}) &= p(\mathbf{x}_{N+1}|\mathbf{z}_{N+1}) \\
p(\mathbf{z}_{N+1}|\mathbf{z}_N,\mathbf{X}) &= p(\mathbf{z}_{N+1}|\mathbf{z}_N)
\end{aligned}
$$

Algorithms – Forward-Backward

Using Bayes' theorem, and we have:

$$\gamma(\mathbf{z}_n) = p(\mathbf{z}_n|\mathbf{X}) = \frac{p(\mathbf{X}|\mathbf{z}_n)p(\mathbf{z}_n)}{p(\mathbf{X})}$$

$$\gamma(\mathbf{z}_n) = \frac{p(\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{z}_n)p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n)}{p(\mathbf{X})} = \frac{\alpha(\mathbf{z}_n)\beta(\mathbf{z}_n)}{p(\mathbf{X})}$$

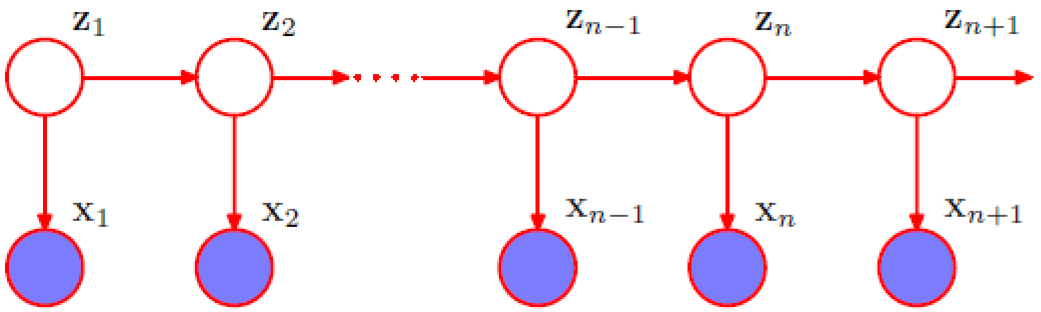
Where we defined:

$$\begin{aligned}\alpha(\mathbf{z}_n) &\equiv p(\mathbf{x}_1,\ldots,\mathbf{x}_n,\mathbf{z}_n)\\ \beta(\mathbf{z}_n) &\equiv p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n)\end{aligned}$$

浙江大学
Zhejiang University

Algorithms – Forward-Backward

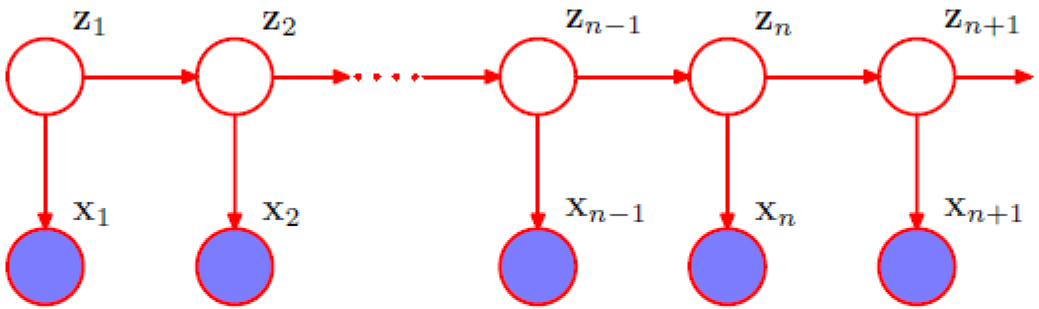
$$
\begin{aligned}
\alpha(\mathbf{z}_n) &= p(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{z}_n) \\
&= p(\mathbf{x}_1, \ldots, \mathbf{x}_n | \mathbf{z}_n) p(\mathbf{z}_n) \\
&= p(\mathbf{x}_n | \mathbf{z}_n) p(\mathbf{x}_1, \ldots, \mathbf{x}_{n-1} | \mathbf{z}_n) p(\mathbf{z}_n) \\
&= p(\mathbf{x}_n | \mathbf{z}_n) p(\mathbf{x}_1, \ldots, \mathbf{x}_{n-1}, \mathbf{z}_n) \\
&= p(\mathbf{x}_n | \mathbf{z}_n) \sum_{\mathbf{z}_{n-1}} p(\mathbf{x}_1, \ldots, \mathbf{x}_{n-1}, \mathbf{z}_{n-1}, \mathbf{z}_n) \\
&= p(\mathbf{x}_n | \mathbf{z}_n) \sum_{\mathbf{z}_{n-1}} p(\mathbf{x}_1, \ldots, \mathbf{x}_{n-1}, \mathbf{z}_n | \mathbf{z}_{n-1}) p(\mathbf{z}_{n-1}) \\
&= p(\mathbf{x}_n | \mathbf{z}_n) \sum_{\mathbf{z}_{n-1}} p(\mathbf{x}_1, \ldots, \mathbf{x}_{n-1} | \mathbf{z}_{n-1}) p(\mathbf{z}_n | \mathbf{z}_{n-1}) p(\mathbf{z}_{n-1}) \\
&= p(\mathbf{x}_n | \mathbf{z}_n) \sum_{\mathbf{z}_{n-1}} p(\mathbf{x}_1, \ldots, \mathbf{x}_{n-1}, \mathbf{z}_{n-1}) p(\mathbf{z}_n | \mathbf{z}_{n-1})
\end{aligned}
$$



Then we have: $\alpha(\mathbf{z}_n) = p(\mathbf{x}_n | \mathbf{z}_n) \sum_{\mathbf{z}_{n-1}} \alpha(\mathbf{z}_{n-1}) p(\mathbf{z}_n | \mathbf{z}_{n-1})$

Zhejiang University

Algorithms – Forward-Backward



$$\begin{aligned}
\beta(\mathbf{z}_n) &= p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n) \\
&= \sum_{\mathbf{z}_{n+1}} p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N,\mathbf{z}_{n+1}|\mathbf{z}_n) \\
&= \sum_{\mathbf{z}_{n+1}} p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n,\mathbf{z}_{n+1})p(\mathbf{z}_{n+1}|\mathbf{z}_n) \\
&= \sum_{\mathbf{z}_{n+1}} p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_{n+1})p(\mathbf{z}_{n+1}|\mathbf{z}_n) \\
&= \sum_{\mathbf{z}_{n+1}} p(\mathbf{x}_{n+2},\ldots,\mathbf{x}_N|\mathbf{z}_{n+1})p(\mathbf{x}_{n+1}|\mathbf{z}_{n+1})p(\mathbf{z}_{n+1}|\mathbf{z}_n)
\end{aligned}$$

Then we have: $$\beta(\mathbf{z}_n) = \sum_{\mathbf{z}_{n+1}} \beta(\mathbf{z}_{n+1})p(\mathbf{x}_{n+1}|\mathbf{z}_{n+1})p(\mathbf{z}_{n+1}|\mathbf{z}_n)$$

浙江大學 Zhejiang University

Algorithms – Forward-Backward

Then we estimate $\xi(z_{n-1,j}, z_{nk})$

$$\xi(\mathbf{z}_{n-1}, \mathbf{z}_n) = p(\mathbf{z}_{n-1}, \mathbf{z}_n | \mathbf{X})$$
$$= \frac{p(\mathbf{X}|\mathbf{z}_{n-1}, \mathbf{z}_n)p(\mathbf{z}_{n-1}, \mathbf{z}_n)}{p(\mathbf{X})}$$
$$= \frac{p(\mathbf{x}_1, \ldots, \mathbf{x}_{n-1}|\mathbf{z}_{n-1})p(\mathbf{x}_n|\mathbf{z}_n)p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N|\mathbf{z}_n)p(\mathbf{z}_n|\mathbf{z}_{n-1})p(\mathbf{z}_{n-1})}{p(\mathbf{X})}$$
$$= \frac{\alpha(\mathbf{z}_{n-1})p(\mathbf{x}_n|\mathbf{z}_n)p(\mathbf{z}_n|\mathbf{z}_{n-1})\beta(\mathbf{z}_n)}{p(\mathbf{X})} \tag{13.43}$$

Then $\xi(z_{n-1,j}, z_{nk})$ can be estimated by α and β recursion.

浙江大学 Zhejiang University

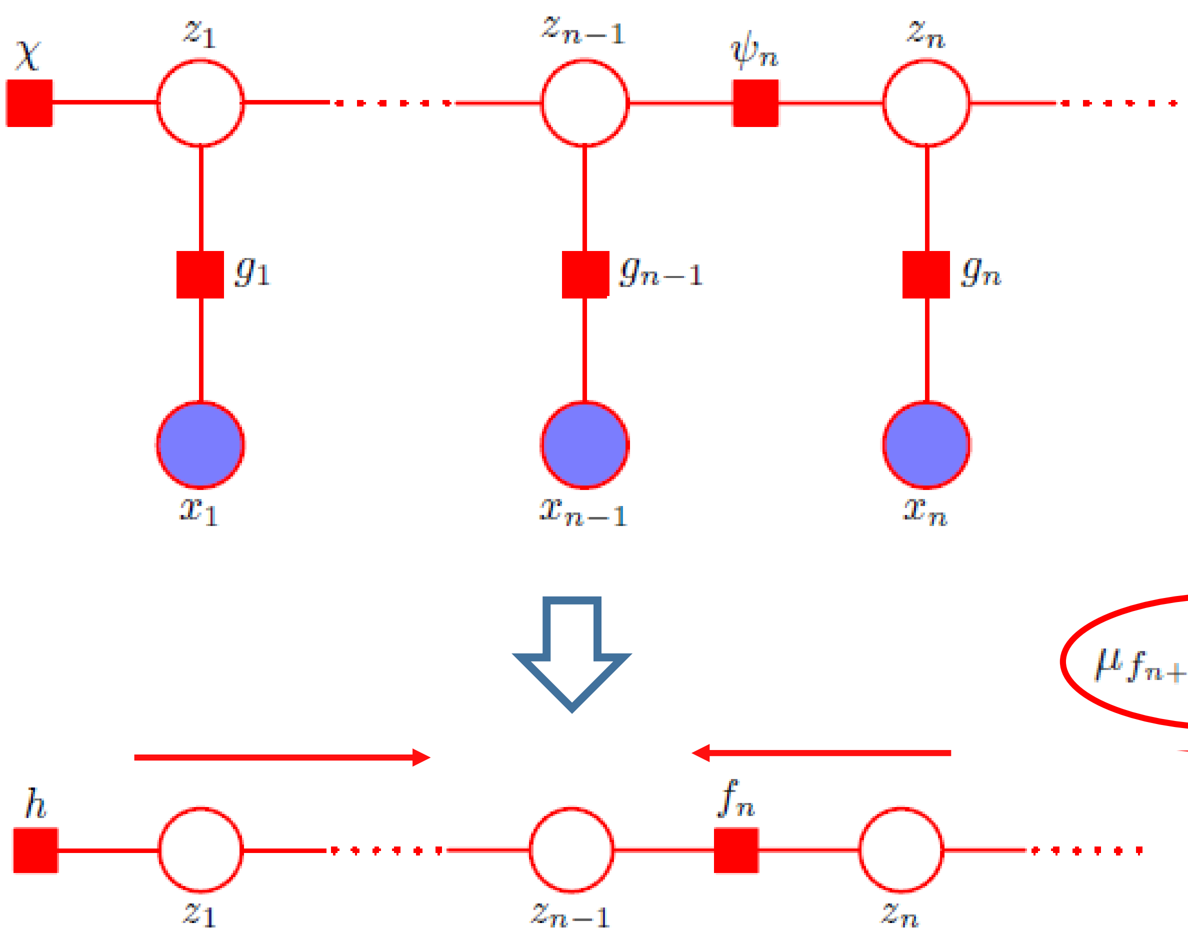
Algorithms – Forward-Backward

For prediction, we get:

$$
\begin{aligned}
p(\mathbf{x}_{N+1}|\mathbf{X}) &= \sum_{\mathbf{z}_{N+1}} p(\mathbf{x}_{N+1}, \mathbf{z}_{N+1}|\mathbf{X}) \\
&= \sum_{\mathbf{z}_{N+1}} p(\mathbf{x}_{N+1}|\mathbf{z}_{N+1}) p(\mathbf{z}_{N+1}|\mathbf{X}) \\
&= \sum_{\mathbf{z}_{N+1}} p(\mathbf{x}_{N+1}|\mathbf{z}_{N+1}) \sum_{\mathbf{z}_N} p(\mathbf{z}_{N+1}, \mathbf{z}_N|\mathbf{X}) \\
&= \sum_{\mathbf{z}_{N+1}} p(\mathbf{x}_{N+1}|\mathbf{z}_{N+1}) \sum_{\mathbf{z}_N} p(\mathbf{z}_{N+1}|\mathbf{z}_N) p(\mathbf{z}_N|\mathbf{X}) \\
&= \sum_{\mathbf{z}_{N+1}} p(\mathbf{x}_{N+1}|\mathbf{z}_{N+1}) \sum_{\mathbf{z}_N} p(\mathbf{z}_{N+1}|\mathbf{z}_N) \frac{p(\mathbf{z}_N, \mathbf{X})}{p(\mathbf{X})} \\
&= \frac{1}{p(\mathbf{X})} \sum_{\mathbf{z}_{N+1}} p(\mathbf{x}_{N+1}|\mathbf{z}_{N+1}) \sum_{\mathbf{z}_N} p(\mathbf{z}_{N+1}|\mathbf{z}_N) \alpha(\mathbf{z}_N)
\end{aligned}
$$

Algorithms – Sum-Product

In the perspective of belief propagation:



$$\begin{aligned} h(\mathbf{z}_1) &= p(\mathbf{z}_1)p(\mathbf{x}_1|\mathbf{z}_1) \\ f_n(\mathbf{z}_{n-1}, \mathbf{z}_n) &= p(\mathbf{z}_n|\mathbf{z}_{n-1})p(\mathbf{x}_n|\mathbf{z}_n) \end{aligned}$$

$\beta(\mathbf{z}_n)$

$$\mu_{f_{n+1}\to f_n}(\mathbf{z}_n) = \sum_{\mathbf{z}_{n+1}} f_{n+1}(\mathbf{z}_n, \mathbf{z}_{n+1})\mu_{f_{n+2}\to f_{n+1}}(\mathbf{z}_{n+1})$$

$\alpha(\mathbf{z}_n)$

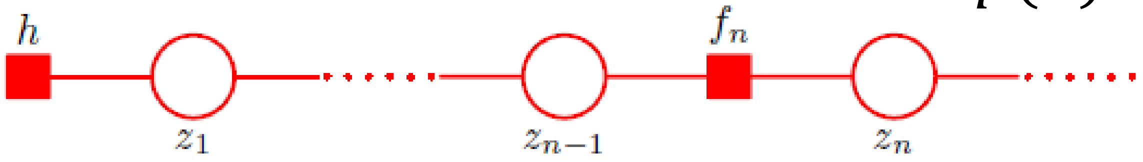
Algorithms – Sum-Product

Estimation of $\gamma(\mathbf{z}_n)$ and $\xi(\mathbf{z}_{n-1,}, \mathbf{z}_n)$:

$$p(\mathbf{z}_n, \mathbf{X}) = \mu_{f_n \to \mathbf{z}_n}(\mathbf{z}_n)\mu_{f_{n+1} \to \mathbf{z}_n}(\mathbf{z}_n) = \alpha(\mathbf{z}_n)\beta(\mathbf{z}_n)$$

$$\gamma(\mathbf{z}_n) = \frac{p(\mathbf{z}_n, \mathbf{X})}{p(\mathbf{X})} = \frac{\alpha(\mathbf{z}_n)\beta(\mathbf{z}_n)}{p(\mathbf{X})}$$

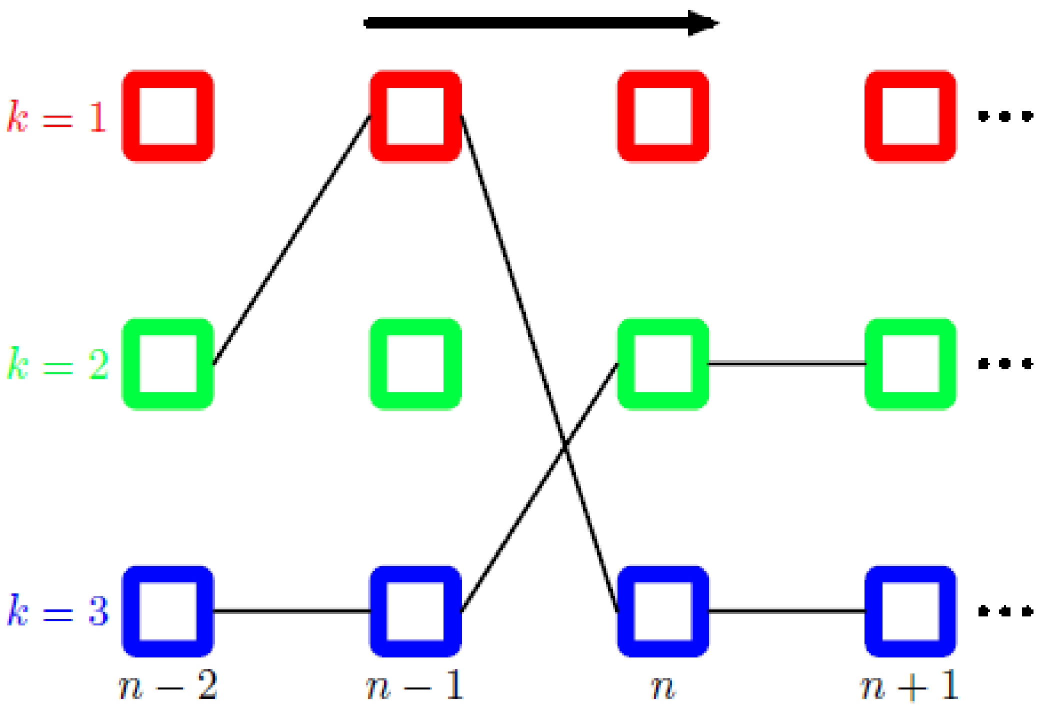
$$\begin{aligned}
\xi(\mathbf{z}_{n-1,}, \mathbf{z}_n) &= p(\mathbf{z}_{n-1,}, \mathbf{z}_n | \mathbf{X}) \\
&= f(\mathbf{z}_{n-1,}, \mathbf{z}_n)\mu_{\mathbf{z}_n \to f_n}(\mathbf{z}_n)\mu_{\mathbf{z}_{n-1} \to f_n}(\mathbf{z}_{n-1}) \\
&= f(\mathbf{z}_{n-1,}, \mathbf{z}_n)\mu_{f_n \to \mathbf{z}_n}(\mathbf{z}_n)\mu_{f_{n-1} \to \mathbf{z}_{n-1}}(\mathbf{z}_{n-1}) \\
&= f(\mathbf{z}_{n-1,}, \mathbf{z}_n)\alpha(\mathbf{z}_{n-1})\beta(\mathbf{z}_n) \\
&= \frac{\alpha(\mathbf{z}_{n-1})\beta(\mathbf{z}_n)p(\mathbf{z}_n|\mathbf{z}_{n-1})p(\mathbf{x}_n|\mathbf{z}_n)}{p(\mathbf{X})}
\end{aligned}$$



浙江大学 Zhejiang University

Algorithms – Viterbi

In many applications of hidden Markov models, the latent variables have some meaningful interpretation, and so it is often of interest to find the most probable sequence of hidden states for a given observation sequence.



Algorithms – Viterbi

$$\mu_{\mathbf{z}_n \to f_{n+1}}(\mathbf{z}_n) = \mu_{f_n \to \mathbf{z}_n}(\mathbf{z}_n)$$

$$\mu_{f_{n+1} \to \mathbf{z}_{n+1}}(\mathbf{z}_{n+1}) = \max_{\mathbf{z}_n} \left\{ \ln f_{n+1}(\mathbf{z}_n, \mathbf{z}_{n+1}) + \mu_{\mathbf{z}_n \to f_{n+1}}(\mathbf{z}_n) \right\}$$

$$\omega(\mathbf{z}_{n+1}) = \ln p(\mathbf{x}_{n+1}|\mathbf{z}_{n+1}) + \max_{\mathbf{z}_n}\{\ln p(\mathbf{z}_{n+1}|\mathbf{z}_n) + \omega(\mathbf{z}_n)\}$$

$$\omega(\mathbf{z}_n) \equiv \mu_{f_n \to \mathbf{z}_n}(\mathbf{z}_n)$$

Where

$$\omega(\mathbf{z}_1) = \ln p(\mathbf{z}_1) + \ln p(\mathbf{x}_1|\mathbf{z}_1)$$

By taking the logarithm and then exchanging maximizations and summations,

$$\omega(\mathbf{z}_n) = \max_{\mathbf{z}_1, \ldots, \mathbf{z}_{n-1}} p(\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{z}_1, \ldots, \mathbf{z}_n)$$

浙江大学 Zhejiang University

Contents

浙江大学
Zhejiang University

Application – HMM in KG

问题：给定训练好的模型，给定一句话，预测每个词对应的实体标签

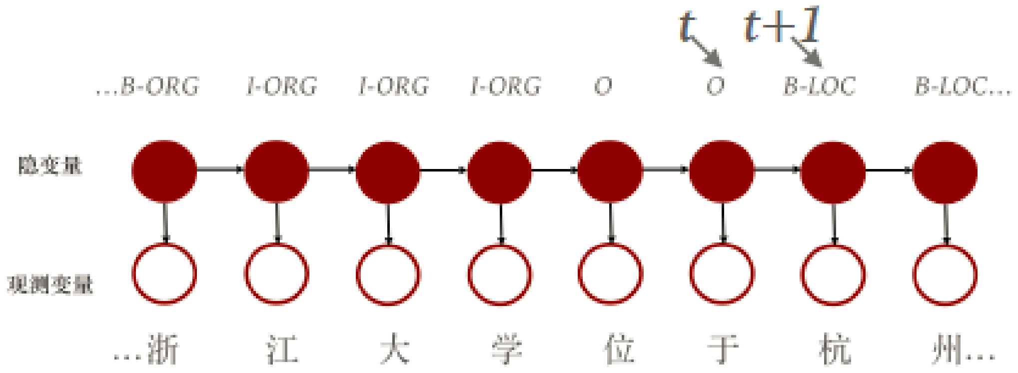
输入：模型$\lambda=(A,B,\Pi)$，观测序列$O=$(浙，江，大，学，位，于，杭，州)
输出：最有可能的隐藏状态序列$I=\{i_1,i_2,\dots i_T\}$，即实体标签序列

动态规划算法的局部状态：在时刻t隐藏状态为i所有可能的状态转移路径$i1,i2,\dots it$中的概率最大值

$$\delta_t(i)=\max_{i_1,i_2,\dots i_{t-1}} P(i_t=i,i_1,i_2,\dots i_{t-1},o_t,o_{t-1},\dots o_1|\lambda),\ \ i=1,2,\dots N$$

核心递推式：

$$\begin{aligned}\delta_{t+1}(i)&=\max_{i_1,i_2,\dots i_t} P(i_{t+1}=i,i_1,i_2,\dots i_t,o_{t+1},o_t,\dots o_1|\lambda)\\&=\max_{1\le j\le N}[\delta_t(j)a_{ji}]b_i(o_{t+1})\end{aligned}$$



Application – HMM in KG

Zhejiang University

问题：给定训练好的模型，给定一句话，预测每个词对应的实体标签

输入：模型$\lambda=(A,B,\Pi)$，观测序列$O=$(浙，江，大，学，位，于，杭，州)
输出：最有可能的隐藏状态序列$I=\{i_1,i_2,\ldots i_T\}$，即实体标签序列

1. 初始化局部状态

$$\delta_1(i)=\pi_i b_i(o_1),\ i=1,2\ldots N$$
$$\Psi_1(i)=0,\ i=1,2\ldots N$$

时刻1，输出为o_1时，各个隐藏状态的可能性。

2.进行动态规划递推时刻$t=2,3,\ldots T$时刻的局部状态

在t时刻，所有从t-1时刻的状态j中，取最大概率。

$$\delta_t(i)=\max_{1\le j\le N}[\delta_{t-1}(j)a_{ji}]b_i(0_t),\ i=1,2\ldots N$$
$$\Psi_t(i)=arg\max_{1\le j\le N}[\delta_{t-1}(j)a_{ji}],\ i=1,2\ldots N$$

从t-1时刻的状态中，选择使t时刻概率最大的那个隐藏状态的编号

3. 如此递推，可计算最后时刻T最大的$\delta_T(i)$,即为最可能隐藏状态序列出现的概率

$$P*=\max_{1\le j\le N}\delta_T(i)$$

4. 计算时刻T最大的$\Psi_t(i)$,即为时刻T最可能的隐藏状态。 $i_T^*=arg\max_{1\le j\le N}[\delta_T(i)]$

5. 利用局部状态$\Psi(i)$开始回溯，最终得到解码的序列，如：“ …B-ORG，I-ORG，I-ORG，I-ORG，O，O，B-LOC，B-LOC…”。

Application

- Text summarization[1]
- Named Entity Recognition[2]
- Spectral Algorithm[3]
- …

[1]Conroy J M, O'leary D P. Text summarization via hidden markov models[C]//Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2001: 406-407.
[2] Morwal S, Jahan N, Chopra D. Named entity recognition using hidden Markov model (HMM)[J]. International Journal on Natural Language Computing (IJNLC), 2012, 1(4): 15-23.
[3] Hsu D, Kakade S M, Zhang T. A spectral algorithm for learning hidden Markov models[J]. Journal of Computer and System Sciences, 2012, 78(5): 1460-1480.

Contents

浙江大學
Zhejiang University

Conclusion

- A method for learning sequential observations
- An Extraordinary way to extract latent features
- A fascinating example for belief propagation

浙江大學
Zhejiang University