法律声明

- 课程详情请咨询
  - 微信公众号：小象
  - 新浪微博：ChinaHadoop

HMM/CRF

主讲人：　史兴

07/28/2017

提纲

- POS tagging / 命名实体识别
- Hidden Markov Model (HMM)
- Conditional Random Field (CRF)
- “Seq2Seq+”
- 课程总结以及展望

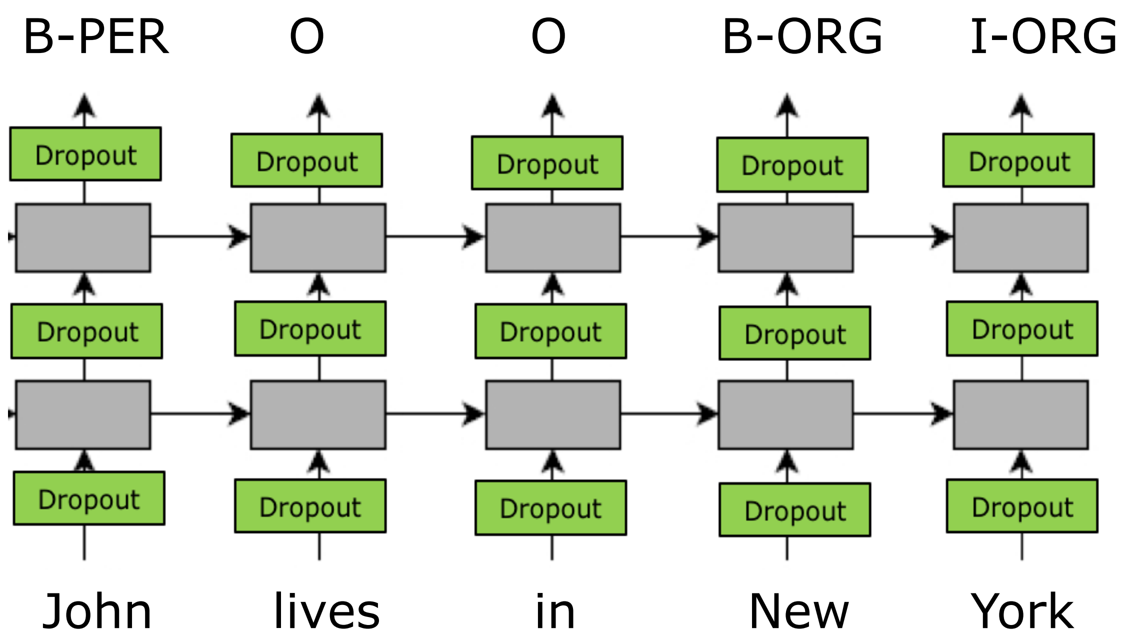
POS tagging / 命名实体识别

- 序列标注问题
  - $x_1, \dots, x_n \Rightarrow y_1, \dots, y_n$
  - x和y一一对应(翻译问题？)
- Part-of-speech Tagging 词性标注
  - I like you => PRON VERB PRON
- 命名实体识别
  - BIO标注

John	lives	in	New	York	and	works	for	the	European	Union
B-PER	O	O	B-LOC	I-LOC	O	O	O	O	B-ORG	I-ORG

POS tagging / 命名实体识别

- 序列标注问题
  - $x_1, \ldots, x_n \Rightarrow y_1, \ldots, y_n$
- 使用Sequence模型解决 + beam search

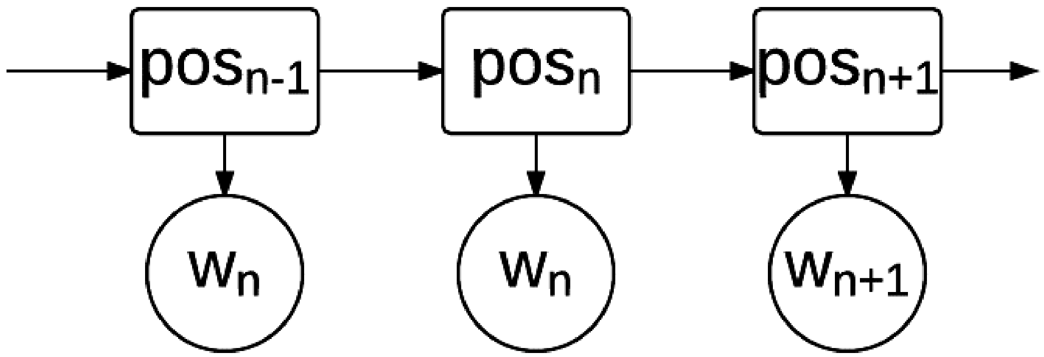


Hidden Markov Model

- Hidden Markov Model (HMM)
- 数据形式
  - 隐含状态: $y_1, \ldots, y_n$
  - 观察到的状态: $x_1, \ldots, x_n$
- 计算联合概率：
  - $p(x_1^n, y_1^n) = p(y_1)p(y_2|y_1)\ldots p(y_n|y_{n-1}) * \mathrm{p}(\mathrm{x}_1|\mathrm{y}_1)\ldots \mathrm{p}(\mathrm{x}_\mathrm{n}|\mathrm{y}_\mathrm{n})$
- 参数:
  - 状态转移矩阵: $p(y_n|y_{n-1})$: k*k 个数字
  - 输出概率： $\mathrm{p}(\mathrm{x}_\mathrm{n}|\mathrm{y}_\mathrm{n})$: k * |V| 个数字

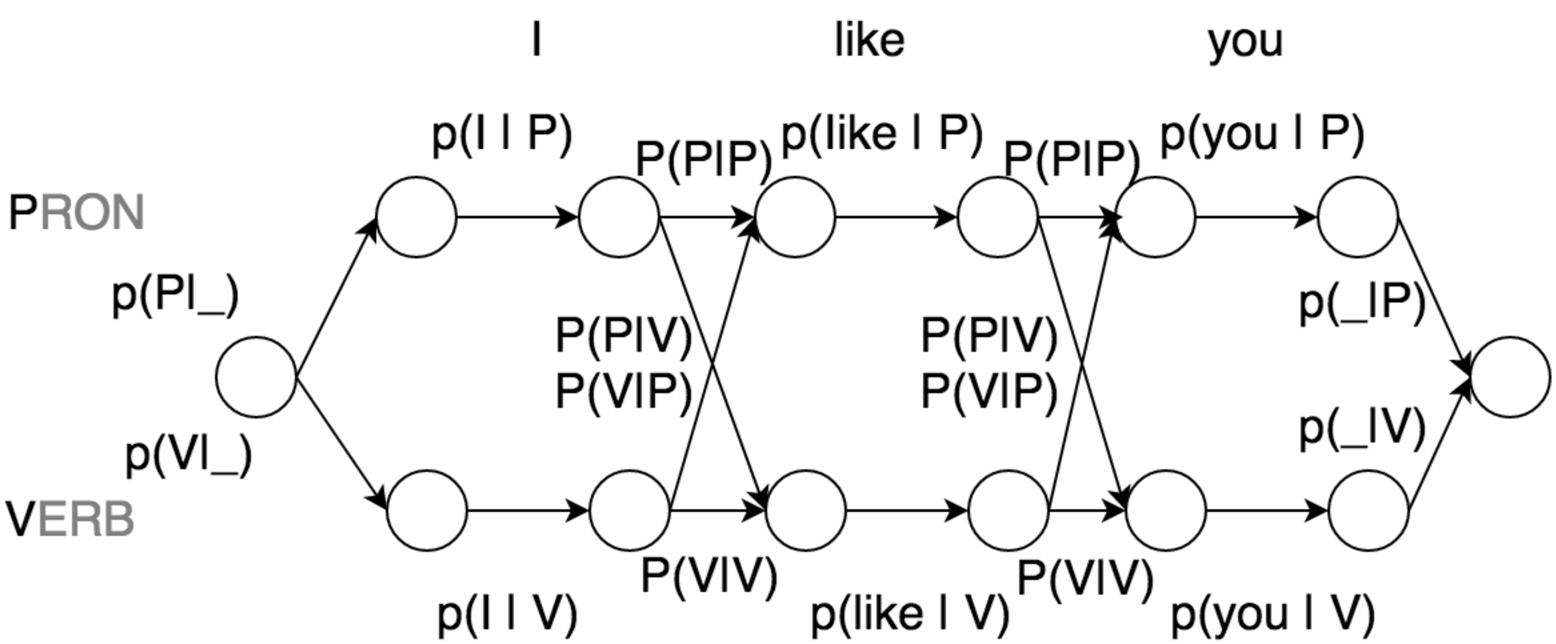
Hidden Markov Model

- Hidden Markov Model
  - 状态转移概率 == FSA
  - 输出概率 == FST
  - HMM == FSA + FST



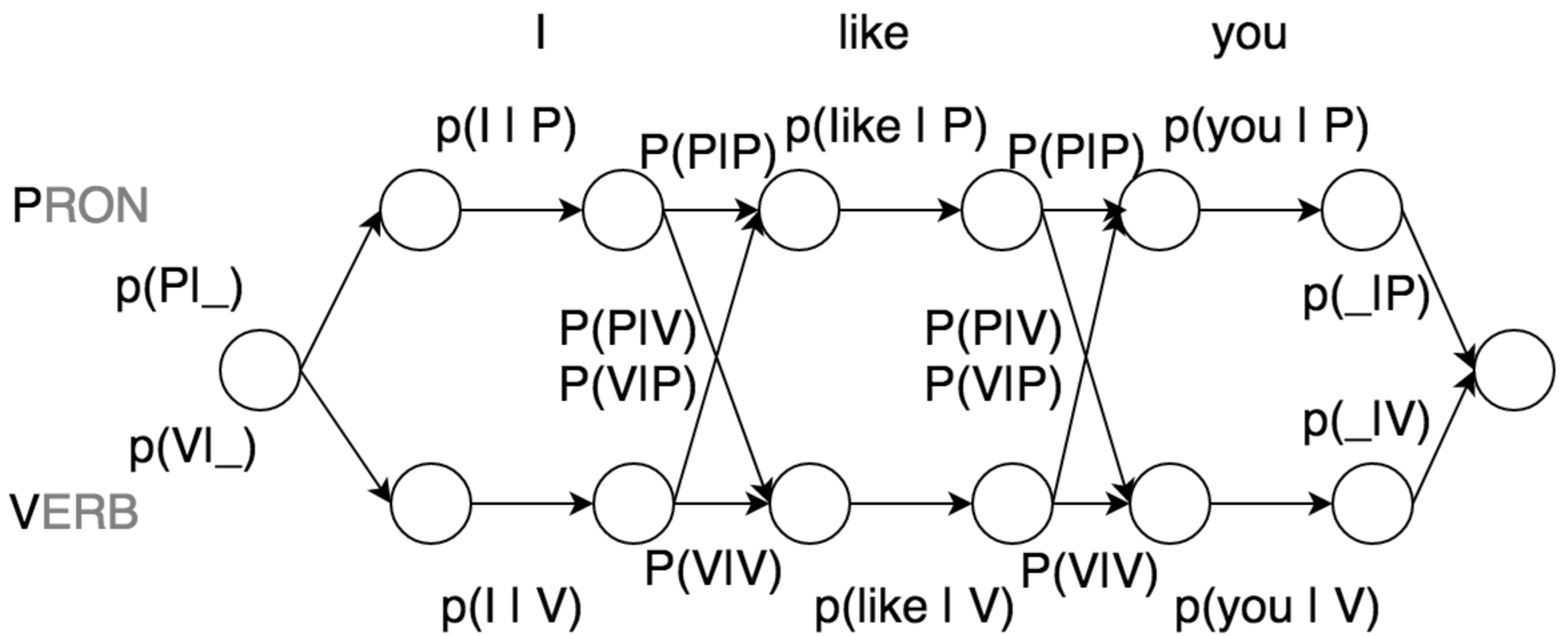
Hidden Markov Model

- Noisy Channel Model和HMM
  - 统一用lattice来表示



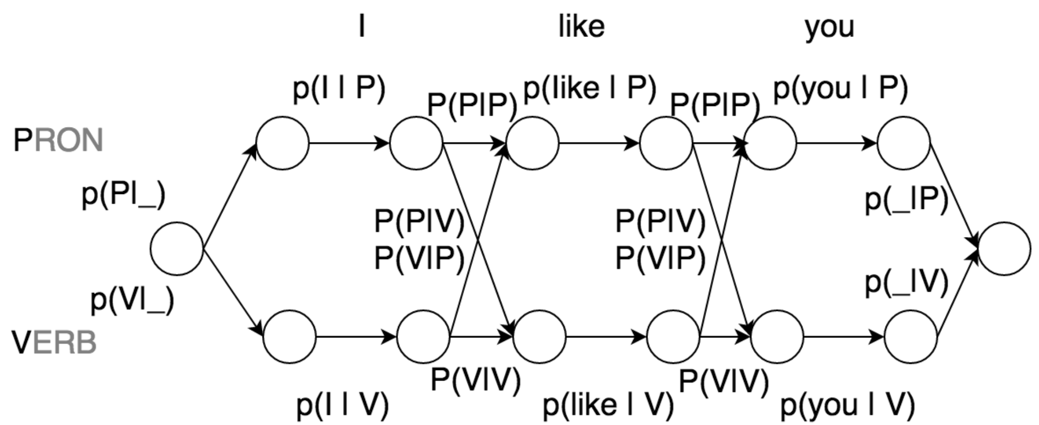
Hidden Markov Model

- 假设参数已经知道
  - 输入 I like you, 如何求得最佳的sequence?
    - Viterbi Algorithm (为什么不用beam search?)



Hidden Markov Model

- 如何求参数?
  - (sentence, POS) 都知道:
    - MLE: max P(POS, sentence)
  - 只有sentence知道:
    - EM 算法: max $P(sentence) = \sum_{all\ possible\ POS} p(sentence|POS_i)P(POS_i)$



Hidden Markov Model

- EM算法
  - $L = \log p(x|\theta) = \log \sum p(x,z|\theta)$
  - $\frac{\partial L}{\partial \theta}$?
  - $L = \log p(x|\theta) = \sum_z q(z) \log p(x|\theta)$

$$= \sum_z q(z) \log \frac{p(x|\theta)}{q(z)} q(z)$$

$$= \sum_z q(z) \log \frac{p(x,z|\theta)}{q(z)} \frac{q(z)}{p(z|x,\theta)}$$

$$= \sum_z q(z) \log \frac{p(x,z|\theta)}{q(z)} + \sum_z q(z) \log \frac{q(z)}{p(z|x,\theta)}$$

Hidden Markov Model

- EM算法
  - $L = \sum_z q(z) \log \frac{p(x,z|\theta)}{q(z)} + \sum_z q(z) \log \frac{q(z)}{p(z|x,\theta)}$

    下界 KL(q||p) >= 0
  - E step: 令 $q(z) = p(z|x,\theta)$, 使得KL = 0
    - 求 $p(z|x,\theta_{old})$
  - M step: 将$p(z|x,\theta_{old})$代入下界中，最大化下界
    - $\theta_{new} = argmax_\theta \sum_z p(z|x,\theta_{old}) \log p(x,z|\theta)$

Hidden Markov Model

- EM算法 手绘理解上界下界

Hidden Markov Model

- HMM中的EM算法
  - 初始化参数： θ_{old} (转移矩阵和输入概率)
  - E step: 令 $q(z) = p(z|x, \theta)$, 使得KL = 0
    - 求 $p(z|x, \theta_{old})$
    - z 是什么？ (POS 序列)
      - z 有 |tag|^n 种
      - 简单的方法，我们来一一枚举这些不同的z
  - 简单的例子：
    - POS tags: “x” 和 “y”
    - 字典： “a” 和 “b”
    - 训练数据： “ a b a”

Hidden Markov Model

- Naïve EM: E step
  - 求 $p(z|x,\theta_{old})$

data completion	P(t1)	P(w1|t1)	P(t2|t1)	P(w2|t2)	P(t3|t2)	P(w3|t3)	P(t,w)	norm P(t,w)
aba → xxx	0.6	0.5	0.6	0.5	0.6	0.5	.027	.216
aba → xxy	0.6	0.5	0.6	0.5	0.4	0.5	.018	.144
aba → xyx	0.6	0.5	0.4	0.5	0.9	0.5	.027	.216
aba → xyy	0.6	0.5	0.4	0.5	0.1	0.5	.003	.024
aba → yxx	0.4	0.5	0.9	0.5	0.6	0.5	.027	.216
aba → yxy	0.4	0.5	0.9	0.5	0.4	0.5	.018	.144
aba → yyx	0.4	0.5	0.1	0.5	0.9	0.5	.0045	.036
aba → yyy	0.4	0.5	0.1	0.5	0.1	0.5	.0005	.004

Hidden Markov Model

- Naïve EM: M step
  - $\theta_{new} = argmax_{\theta} \sum_{z} p(z|x,\theta_{old}) \log p(x,z|\theta)$

Fractional counts

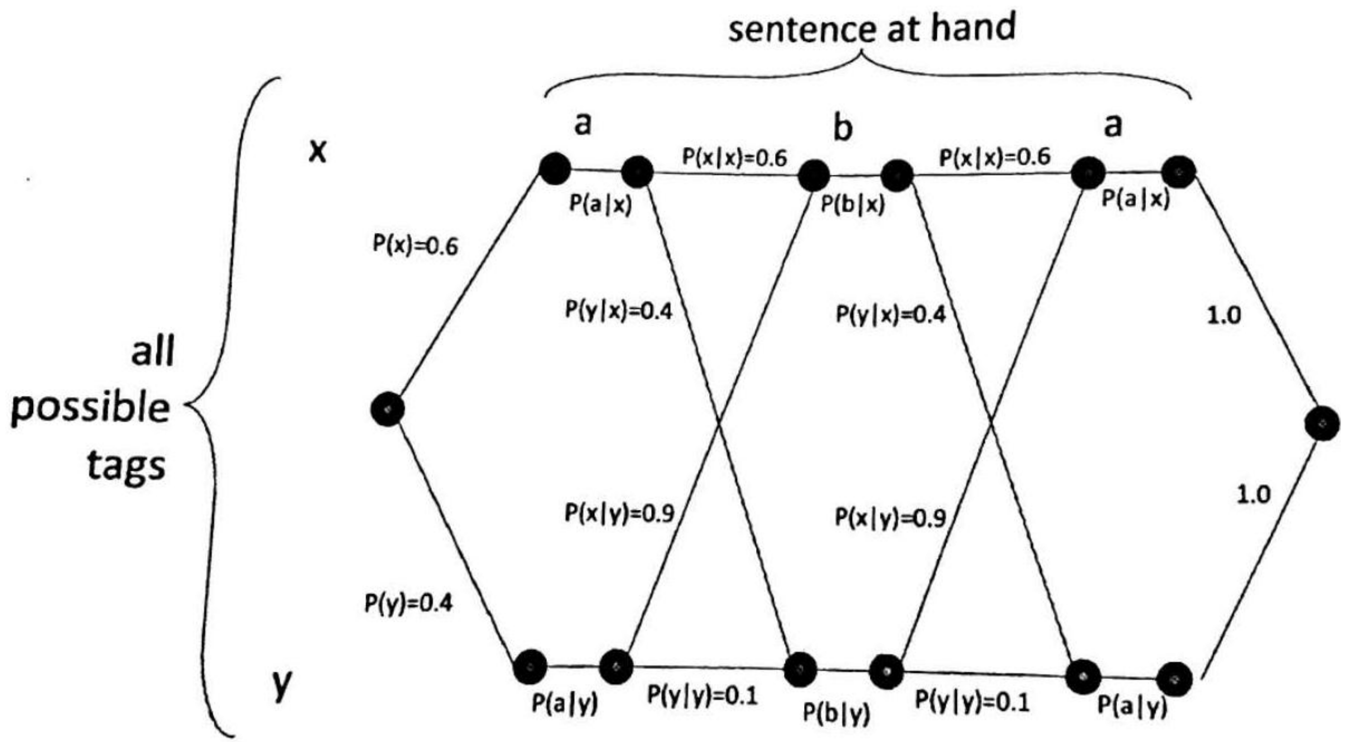
- count(x, a) = .216 * 2 + .144 + .216 * 2 + .024 + .216 + .036 = 1.284
- count(x, b) = .216 + .144 + .216 + .144 = $\boxed{0.72}$ remember this

Revised probability values:

- P(a | x) = 1.284 / (1.284 + 0.72) = 0.64
- P(b | x) = 0.72 / (1.284 + 0.72) = 0.36

Hidden Markov Model

- 使用lattice的E step
  - Forward-backward algorithm

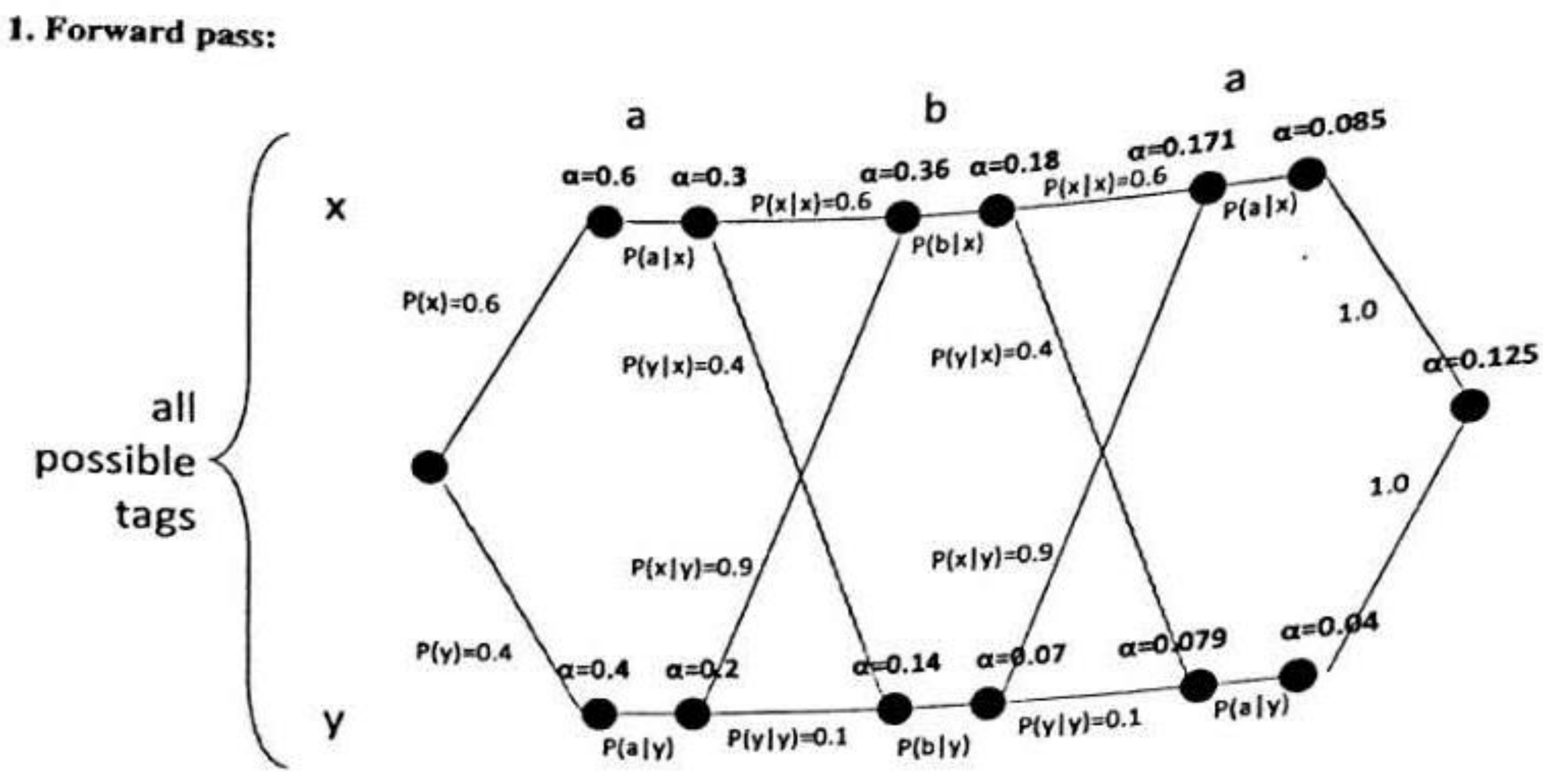


Hidden Markov Model

- 使用lattice的E step
  - Forward-backward algorithm
  - $fc(\theta) = 0$
  - 对训练数据中的每句话，建立一个lattice:
    - 对于每个点：
      - a(node): 从start到node的所有路径上的概率之和
      - b(node): 从node到end的所有路径上的概率之和
      - a(node)和b(node) 都可以用动态规划来计算
    - 对于每条边(m->n), 需要计算其fractional_count
      - fc(m,n) = a(m) * p(m->n) * b(n) / a(end)
    - 每条边都对应着某个参数θ
      - $fc(\theta)$ += fc(m,n)

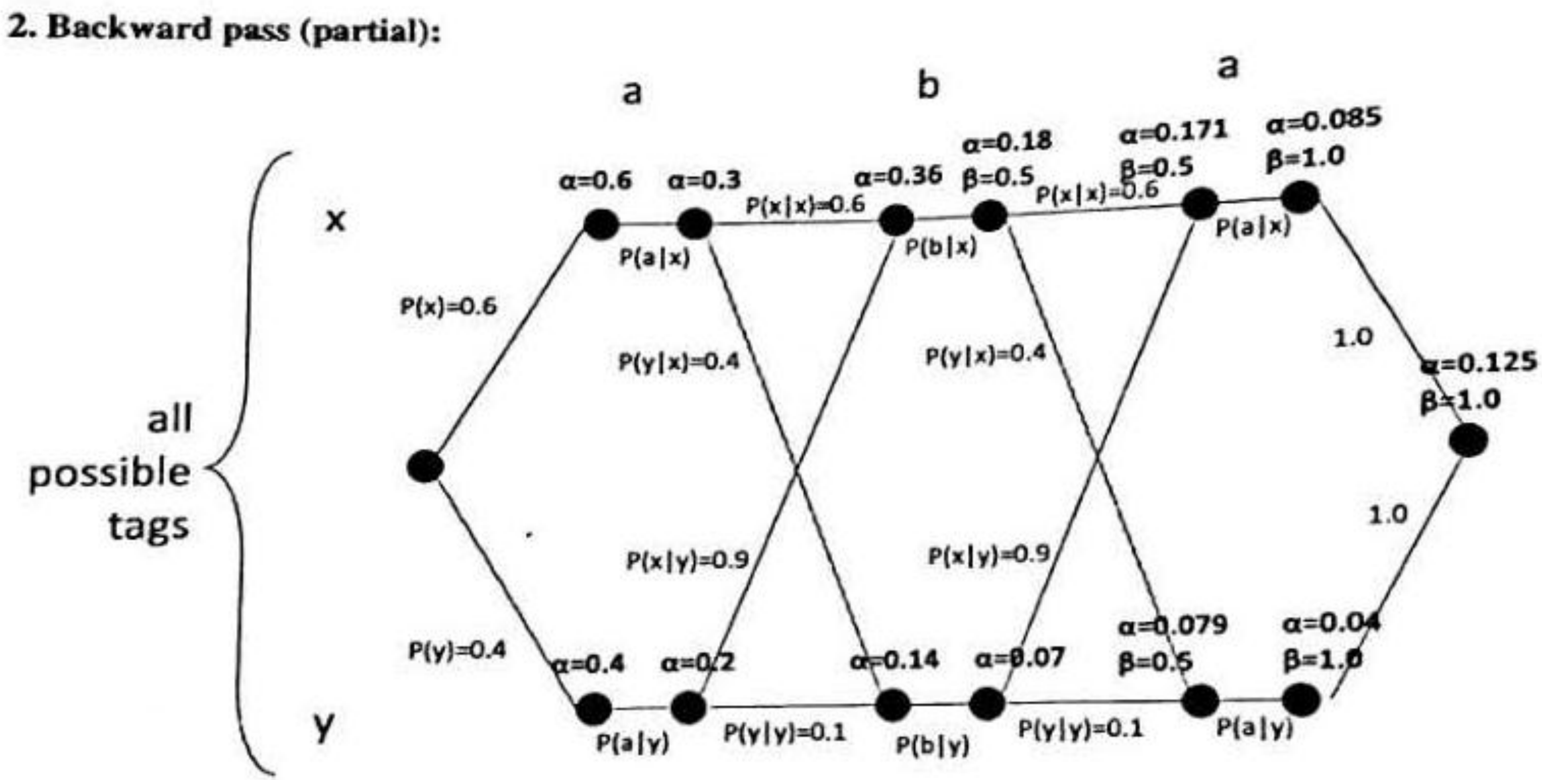
Hidden Markov Model

- 使用lattice的E step



Hidden Markov Model

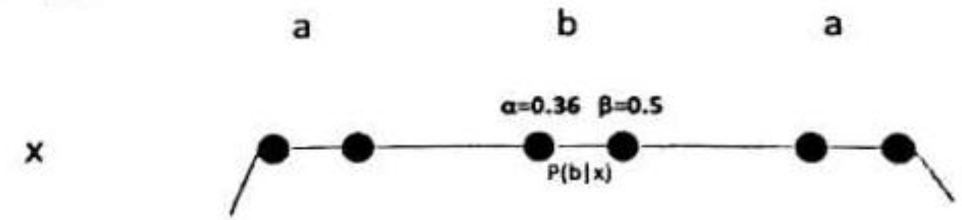
- 使用lattice的E step



Hidden Markov Model

- 使用lattice的E step

3. Count collection (for P(b|x) link only):



count(x, b) += 0.36 * 0.5 * 0.5 / 0.125 = $\boxed{0.72}$

Hidden Markov Model

- 使用lattice的M step
  - fc(θ) normalize, 使其变成概率

Conditional Random Field

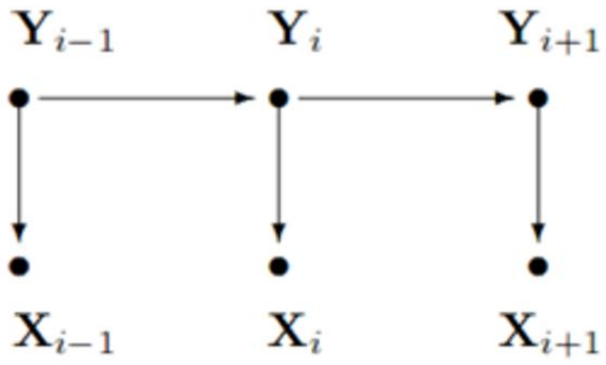
- HMM
  - 计算联合概率：
  - $p(x_1^n, y_1^n) = p(y_1)p(y_2|y_1) \dots p(y_n|y_{n-1}) * \mathrm{p}(\mathrm{x}_1|\mathrm{y}_1) \dots \mathrm{p}(\mathrm{x}_\mathrm{n}|\mathrm{y}_\mathrm{n})$
- Conditional Random Field (CRF)
  - 计算条件概率
  - $p(y_1^n|x_1^n) = \frac{1}{Z} \exp(\sum_{k=1}^n \sum_i \lambda_i f_i(y_{k-1}, y_k, x_1^n, k))$
  - $\mathrm{Z} = \sum_{y_1^n} \exp(\sum_{k=1}^n \sum_i \lambda_i f_i(y_{k-1}, y_k, x_1^n, k))$
  - 参数：λ_i

Conditional Random Field

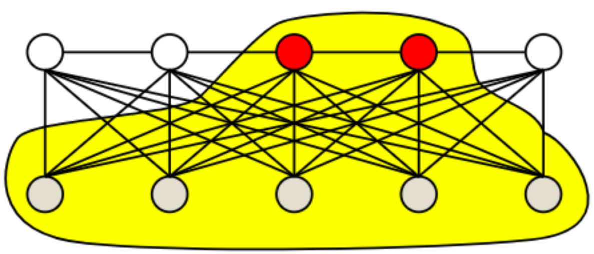
- Conditional Random Field (CRF)
  - $p(y_1^n|x_1^n) = \frac{1}{Z}\exp(\sum_{k=1}^{n}\sum_i \lambda_i f_i(y_{k-1}, y_k, x_1^n, k))$
  - feature function: $f_i(y_{k-1}, y_k, x_1^n, k)$
    - 状态转移：
      - f = 1 if y_{k-1}=B-PER and y_k=I-PER， otherwise 0
    - 输出函数：
      - f = 1 if y_k=B-PER and x_k=John, otherwise 0
    - 更多的feature:
      - f = 1 if y_k=B-PER and x_k的首字母大写
      - f = 1 if y_k=I-PER and x_{k+1}=said

Conditional Random Field

- HMM vs CRF
  - HMM: Directed Graphical Model
  - CRF: Undirected Graphical Model
  - CRF: 可以支持更多的feature
  - HMM: 可以支持无监督学习



HMM



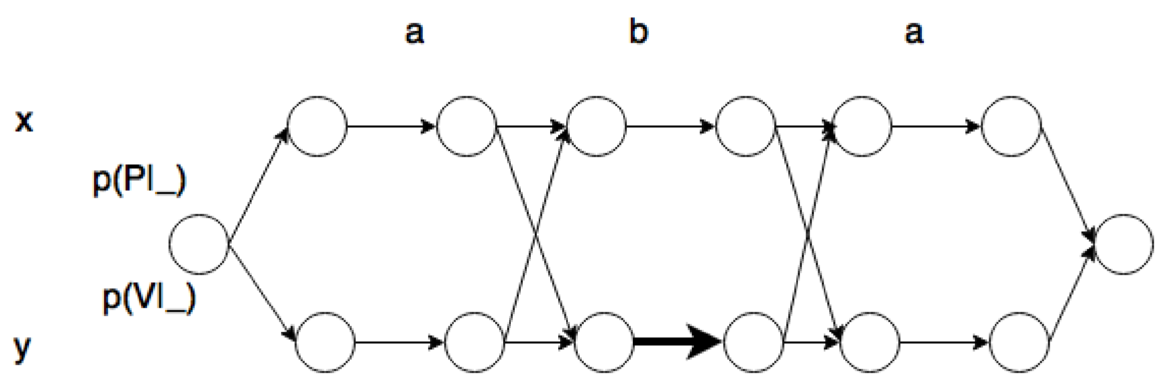
CRF

Conditional Random Field

- Conditional Random Field (CRF)
  - $p(y_1^n|x_1^n) = \frac{1}{Z}\exp(\sum_{k=1}^{n}\sum_i \lambda_i f_i(y_{k-1}, y_k, x_1^n, k))$
  - 如何学习参数λ_i: SGD
    - $\frac{\partial p(y_1^n|x_1^n)}{\partial \lambda_i} = f_i(y_1^n, x_1^n) - \sum_{\widehat{y_1^n}} p(\widehat{y_1^n}|x_1^n) f_i(\widehat{y_1^n}, x_1^n)$
    - $\lambda_i \leftarrow \lambda_i + \eta \frac{\partial p(y_1^n|x_1^n)}{\partial \lambda_i}$
    - 计算$\sum_{\widehat{y_1^n}} p(\widehat{y_1^n}|x_1^n) f_i(\widehat{y_1^n}, x_1^n) = E(f_i)$是难点
      - Naïve : 枚举所有可能的$\widehat{y_1^n}$
      - Forward-backward algorithm

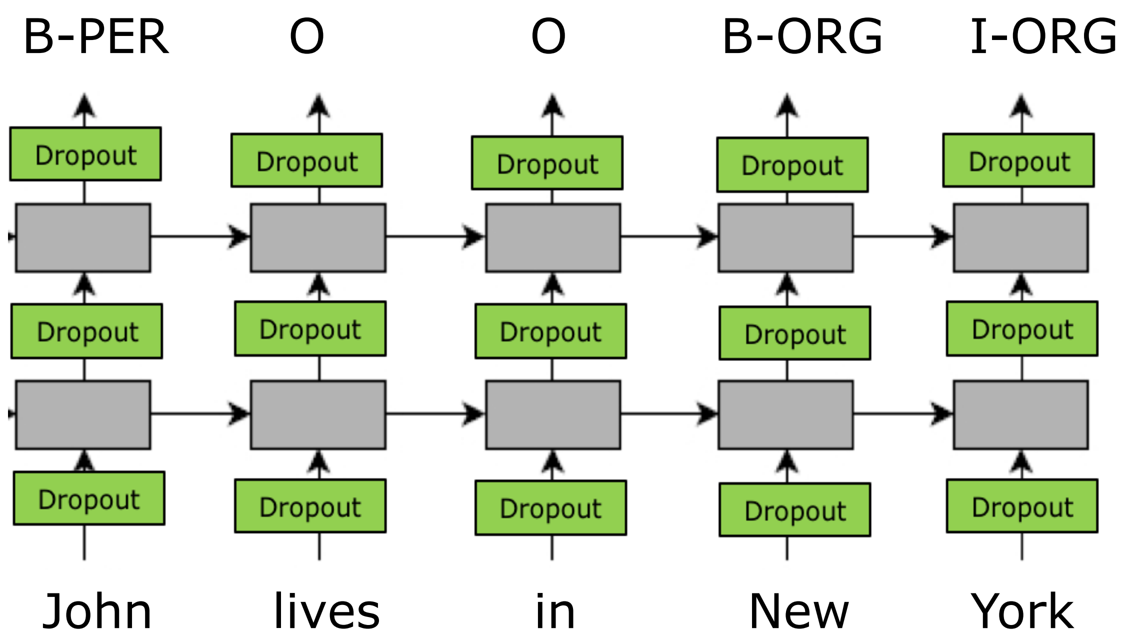
Conditional Random Field

- Conditional Random Field (CRF)
  - 计算$\sum_{\widehat{y_1^n}} p\left(\widehat{y_1^n} \middle| x_1^n\right) f_i\left(\widehat{y_1^n}, x_1^n\right) = \mathrm{E}(\mathrm{f_i})$是难点
    - Forward-backward algorithm
    - score(e) = exp($\sum_i \lambda_i f_i(e)$); e = m→n
    - fc(e) = a(m)∗score(e)∗b(n) / a(end)
    - $E(f_i) = \sum_{e\ in\ lattice} fc(e)\ f_i(e)$



Conditional Random Field

- CRF vs Seq model
  - Seq model: local decision(softmax); beam search
  - CRF: global decision;markov假设,viterbi algorithm

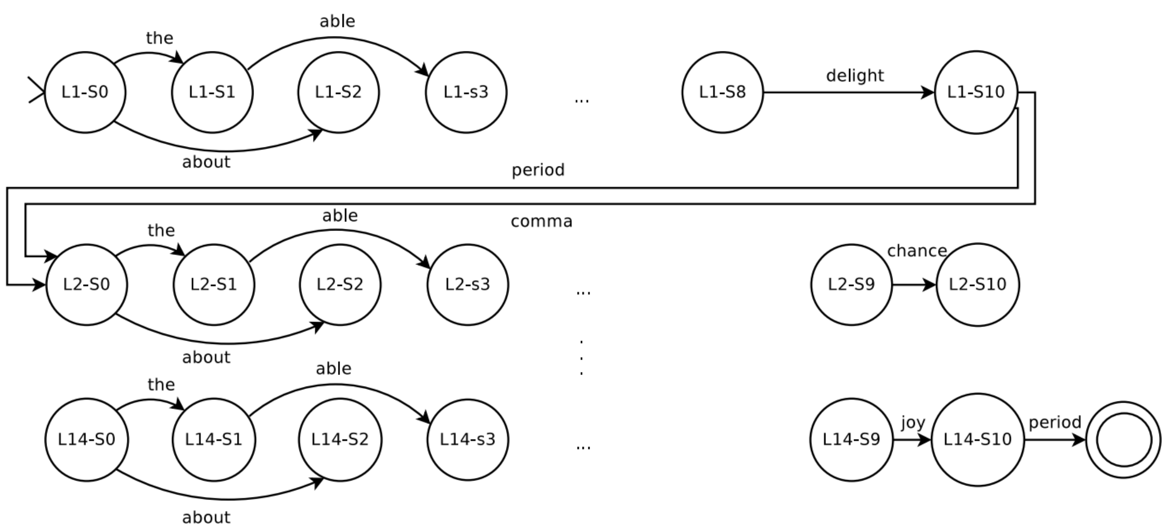


“Seq2Seq+”

- “互联网+”
  - 使用互联网改造传统产业
  - 成功的案例：有互联网思维，同时有对传统业务的经验
- “Seq2Seq+”
  - 使用Seq2Seq模型改造传统模型
  - 成功的案例：有Seq2Seq的思维，同时有对传统模型的经验

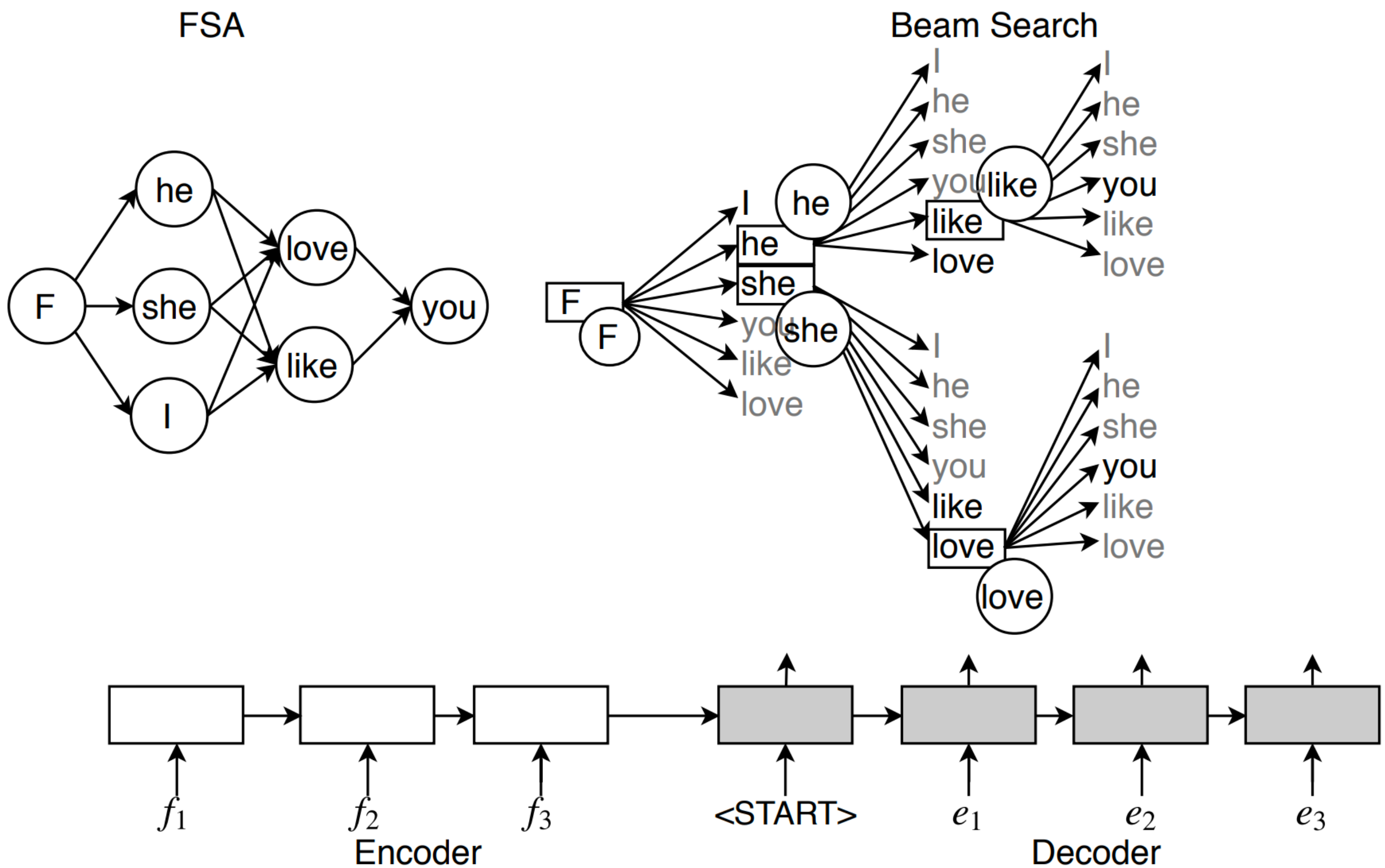
“Seq2Seq+”

- Seq2Seq + FSA
  - 英文诗歌生成：http://52.24.230.241/poem/
  - 14行诗的押韵模式： ABAB CDCD EFEF GG
  - 每一行的重音模式：0101010101



"Seq2Seq+"

- Seq2Seq + FSA



“Seq2Seq+”

- Seq2Seq + FSA
  - 一个自动生成中文rap的程序
    - 有时间，有决心的欢迎联系我

MC Hafez:
(思考了两秒…)

An open space between the gates foundation,
Focus on a new machine translation,
Through the night of mobile penetration,
Had a dream about an innovation.

四押！！

张震岳：我觉得不行，英文太多
MC hotdog：我觉得可以

"Seq2Seq+"

- Seq2Seq +CRF
  - 命名实体识别(NER)的最佳解决方案

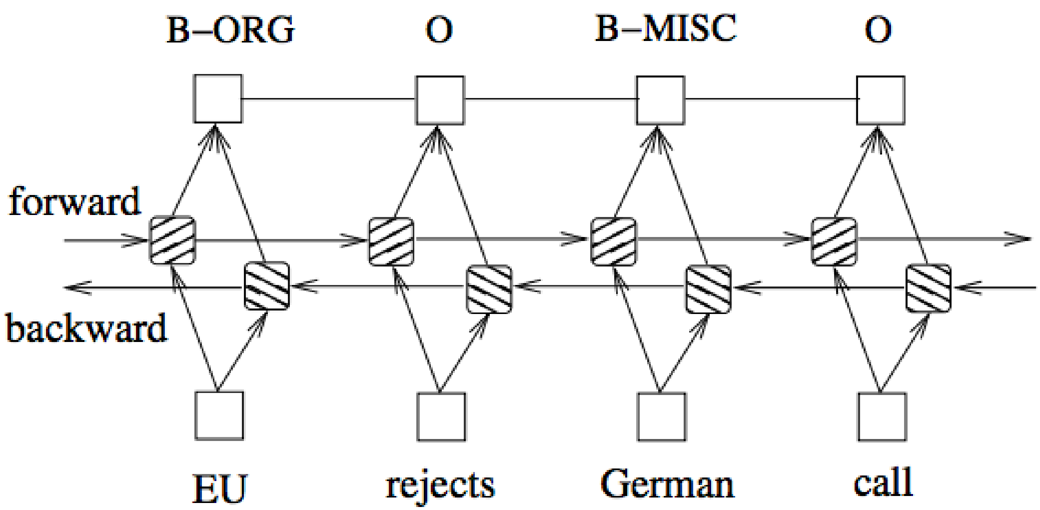


Figure 7: A BI-LSTM-CRF model.

https://arxiv.org/pdf/1508.01991.pdf

课程总结以及展望

- 10节课
  - 介绍课程
  - 语言模型/KenLM
  - 神经网络基础/NNLM/word2vec/超参数搜索
  - Vanilla RNN/LSTM/RNNLM
  - RNN代码详解
  - Seq2Seq/beam search/Attention
  - Seq2Seq/beam search/Attention 代码详解
  - Beam Search 实现细节 / Seq2Seq可视化
  - 自动机
  - HMM/CRF/Seq2Seq+

课程总结以及展望

- 曾经许下的美好的诺言
  - 多种优化方法对比：SGD, Adagrad 等
  - Seq2Seq 模型的提速
    - NCE/LSH/word alignment/Knowledge Distillation
  - 对话生成 / 情感分类 Supervised/Unsupervised
  - 基本分类器：Naive Bayes / Perceptron / SVM / Decision Tree / xgboost

课程总结以及展望

- 期望
  - 至少学会了一样东西
    - LM ?
    - Word2Vec ?
    - RNN ?
    - Seq2Seq ?
    - Beam Search ?
    - Attention ?
    - HMM / CRF / 自动机 ?
  - 今后可以自主的看论文，看代码
  - 学习一下英语，google >? baidu

课程总结以及展望

- 售后服务
  - 欢迎任何问题，请求，建议，批评，合作，推荐，被推荐，咨询，入伙。。。
  - 没有一年的限制

一些CRF的参考链接

https://guillaumegenthial.github.io/sequence-tagging-with-tensorflow.html

http://pages.cs.wisc.edu/~jerryzhu/cs838/CRF.pdf

http://slideplayer.com/slide/4129127/

https://arxiv.org/pdf/1508.01991.pdf

联系我们

小象学院：互联网新技术在线教育领航者

- 微信公众号：大数据分析挖掘
- 新浪微博：ChinaHadoop