

问题

- 共同点
 - 都是用来对观测序列和输出序列建模
- CRF的优势
 - HMM: 最大的缺点是其输出独立性假设, 导致其不能考虑上下文的特征, 限制了特征的选择; 在每一节点都要进行归一化, 所以只能找到局部的最优值, 同时也带来了标记偏见的问题
 - CRF: 选择上下文特征; 不在每个节点进行归一化而是所有特征进行全局归一化, 可以求得全局的最优值

1. 如何理解HMM和朴素贝叶斯、CRF和逻辑斯蒂回归的关系, 以及HMM和CRF的关系? (相同点、不同点)
2. 为什么一般认为CRF比HMM的效果好?

CRFs can define a much larger set of features. Whereas HMMs are necessarily local in nature (because they're constrained to binary transition and emission feature functions, which force each word to depend only on the current label and each label to depend only on the previous label), CRFs can use more global features. For example, one of the features in our POS tagger above increased the probability of labelings that tagged the first word of a sentence as a VERB if the end of the sentence contained a question mark.

CRFs can have arbitrary weights. Whereas the probabilities of an HMM must satisfy certain constraints (e.g., $0 \leq p(w_i | l_i) \leq 1$, $\sum_w p(w_i = w | l_i) = 1$), the weights of a CRF are unrestricted (e.g., $\log p(w_i | l_i)$ can be anything it wants).

CRF既具有判别式模型的优点, 又考虑到长距离上下文标记间的转移概率, 以序列化形式进行**全局参数优化和解码**的特点, 解决了其他判别式模型(如MEMM)难以避免的标记偏见问题。

- 1) HMM是假定满足HMM独立假设。CRF没有, 所以CRF能容纳更多上下文信息。
- 2) CRF计算的是全局最优解, 不是局部最优值。
- 3) CRF是给定观察序列的条件下, 计算整个标记序列的联合概率。而HMM是给定当前状态, 计算下一个状态。
- 4) CRF比较依赖特征的选择和特征函数的格式, 并且训练计算量大

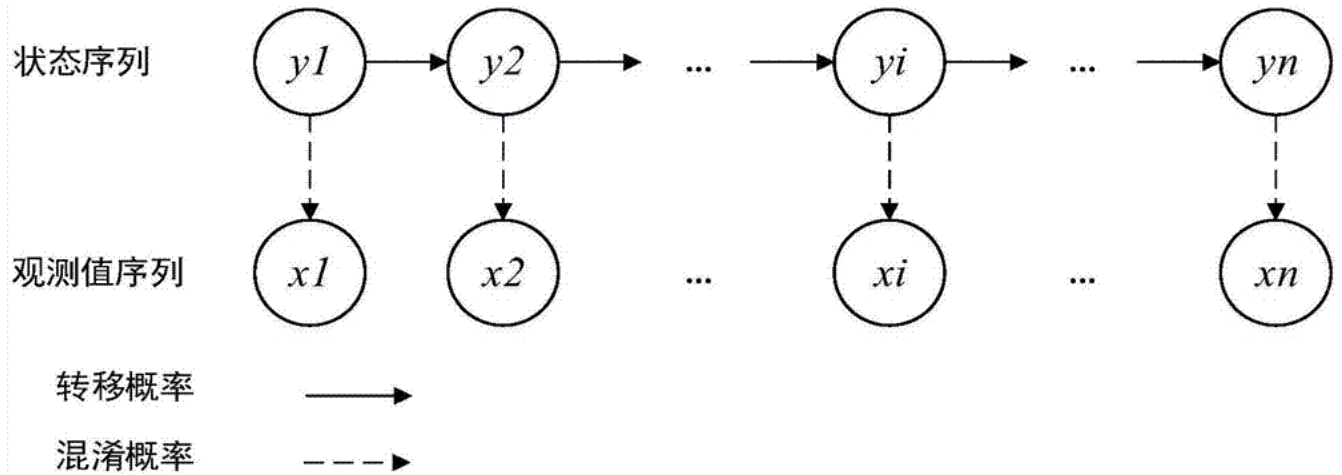
条件随机场 (CRF) 和隐马尔科夫模型 (HMM) 最大区别在哪里? CRF的全局最优体现在哪里? - 尔康的鼻孔的回答 - 知乎 <https://www.zhihu.com/question/53458773/answer/554436625>

如何用简单易懂的例子解释条件随机场 (CRF) 模型? 它和HMM有什么区别? - Scofield的回答 - 知乎 <https://www.zhihu.com/question/35866596/answer/236886066>

HMM & CRF

	图的类型	模型类型	特点 (共同点)	概率计算	预测算法	学习算法
HMM	有向图	生成模型	状态之间的马尔科夫性	前向后向	维特比算法	EM算法
CRF	无向图	判别模型	状态之间的马尔科夫性	前向后向	维特比算法	IIS、拟牛顿法

超全的博客: [HMM](#)



[CRF](#)

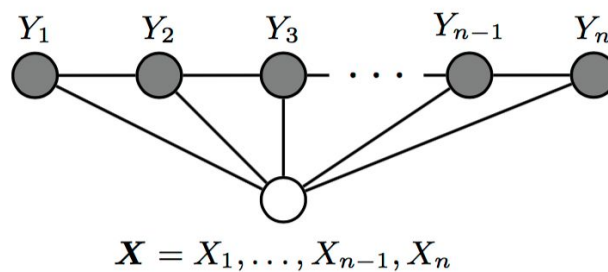
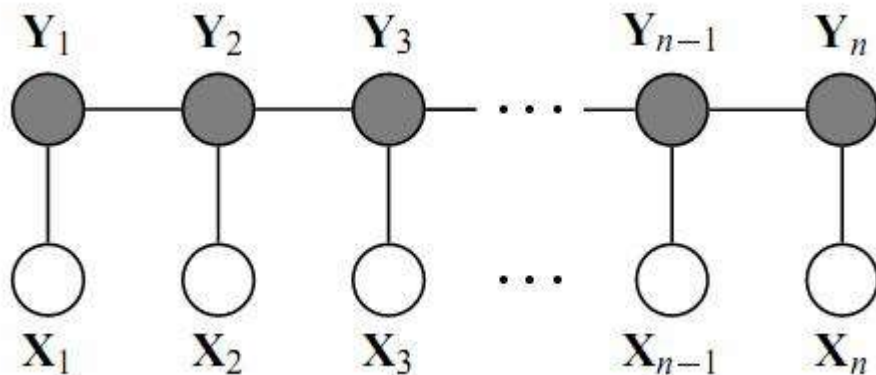


Figure 1: Graphical structure of a chain-structured CRFs for sequences. The variables corresponding to unshaded nodes are *not* generated by the model.



HMM计算的向量化形式

前向

$$\begin{aligned}\alpha_{t+1} &= (A^T \alpha_t) \odot b_{o_t} \\ &= (A \odot b_{o_t})^T \alpha_t\end{aligned}$$

后向

$$\begin{aligned}\beta_t &= (A \beta_{t+1}) \odot b_{o_{t+1}} \\ &= (A \odot b_{o_{t+1}}) \beta_{t+1}\end{aligned}$$

其中, \odot 为element-wise product.

朴素贝叶斯、逻辑斯蒂回归、HMM、CRF的关系

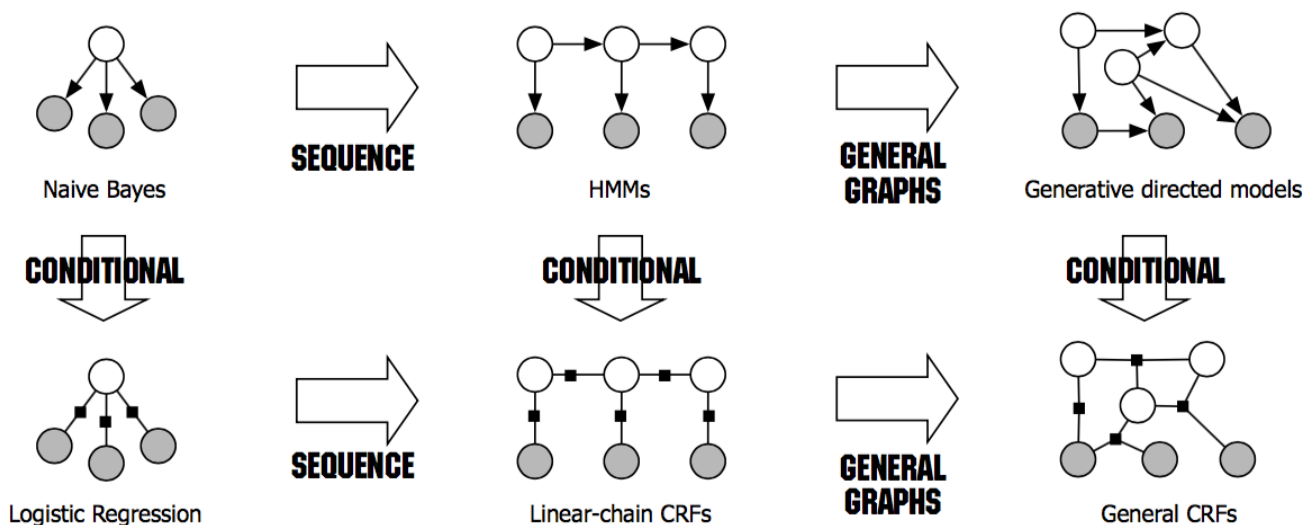
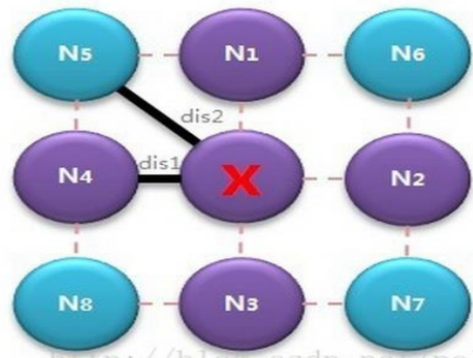


Figure 1.2 Diagram of the relationship between naive Bayes, logistic regression, HMMs, linear-chain CRFs, generative models, and general CRFs.

二维CRF

- 随机场

- 随机场 (Random Field) 中，利用邻域系统可以分析空间上的马尔科夫性。一个像素点的特性，更可能受它周围像素的影响，与它距离越远的像素，对它的特性的影响越小。



- 应用：图像分割