

20180329_条件随机场 隐马模型 和 最大熵隐马模型 的比较

1-

隐马模型HMM：由于其输出独立性假设，导致其不能考虑上下文的特征，限制了特征的选择

最大熵隐马模型MEMM:最大熵隐马模型则解决了这一问题，可以任意的选择特征，但由于其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（label bias），即凡是训练语料中未出现的情况全都忽略掉

条件随机场：条件随机场则很好的解决了这一问题，他并不在每一个节点进行归一化，而是所有特征进行全局归一化，因此可以求得全局的最优值。

条件随机场的训练和开源工具现在只支持链式的序列，复杂的尚不支持。

条件随机场(CRF)相对于HMM,MEMM的优势

首先，CRF，HMM(隐马模型)，MEMM(最大熵隐马模型)都常用来做序列标注的建模，像词性标注，True casing。但隐马模型一个最大的缺点就是**由于其输出独立性假设，导致其不能考虑上下文的特征，限制了特征的选择**，而最大熵隐马模型则解决了这一问题，可以任意的选择特征，但由于其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（label bias），即凡是训练语料中未出现的情况全都忽略掉**，而条件随机场则很好的解决了这一问题，他**并不在每一个节点进行归一化，而是所有特征进行全局归一化，因此可以求得全局的最优值**。目前，条件随机场的训练和解码的开源工具还只支持链式的序列，复杂的尚不支持，而且训练时间很长，但效果还可以。

大致总结一下，详细地用到再看吧:)

MEMM的局限性在于其利用训练的局部模型去做全局预测。其最优预测序列只是通过viterbi算法将局部的最大熵模型结合而成的。另外CRF++是个不错的软件，速度高，性能好，楼主可以试试。

CRF++也只支持链式的，如果想用高阶的feature,可以用pocket crf:

下列哪个不属于CRF模型对于HMM和MEMM模型的优势（ ）

- A.特征灵活
- B.速度快
- C.可容纳较多上下文信息
- D.全局最优

答案：B

analysis

HMM:隐马尔可夫模型

MEMM: 最大熵隐马尔可夫模型

CRF : 条件随机场

这三个模型都可以用来做序列标注模型。但是其各自有自身的特点，HMM模型是对转移概率和表现概率直接建模，统计共现概率。而MEMM模型是对转移 概率和表现概率建立联合概率，统计时统计的是条件概率。MEMM容易陷入局部最优，是因为MEMM只在局部做归一化，而CRF模型中，统计了全局概率，在 做归一化时，考虑了数据在全局的分布，而不是仅仅在局部归一化，这样就解决了MEMM中的标记偏置的问题。

举个例子，对于一个标注任务，“我爱北京天安门”，

标注为"s s b e b c e"

对于HMM的话，其判断这个标注成立的概率为 $P = P(s \text{ 转移到 } s)P('我' \text{ 表现为 } s) P(s \text{ 转移到 } b)P('爱' \text{ 表现为 } s) \dots * P()$ 。训练时，要统计状态转移概率矩阵和表现矩阵。

对于MEMM的话，其判断这个标注成立的概率为 $P = P(s \text{ 转移到 } s | '我' \text{ 表现为 } s)P('我' \text{ 表现为 } s) P(s \text{ 转移到 } b | '爱' \text{ 表现为 } s)P('爱' \text{ 表现为 } s)$ 。训练时，要统计条件状态转移概率矩阵和表现矩阵。

对于CRF的话，其判断这个标注成立的概率为 $P = F(s \text{ 转移到 } s, '我' \text{ 表现为 } s) \dots$ 。F为一个函数，是在全局范围统计归一化的概率而不是像MEMM在局部统计归一化的概率。

优点：

(1) CRF没有HMM那样严格的独立性假设条件，因而可以容纳任意的上下文信息。和HMM相比，特征设计灵活（与ME一样）

(2) 同时，和MEMM相比，由于CRF计算全局最优输出节点的条件概率，它还克服了最大熵马尔可夫模型标记偏置（Label-bias）的缺点。

(3) 和MEMM相比，CRF是在给定需要标记的观察序列的条件下，计算整个标记序列的联合概率分布，而不是在给定当前状态条件下，定义下一个状态的状态分布。

缺点：训练代价大、复杂度高

介绍到这里，就可以得出结论了，如果想了解深入的话，请往下继续看：

HMM模型中存在两个假设：一是输出观察值之间严格独立，二是状态的转移过程中当前状态只与前一状态有关(一阶马尔可夫模型)。

MEMM模型克服了观察值之间严格独立产生的问题，但是由于状态之间的假设理论，使得该模型存在标注偏置问题。

CRF模型解决了标注偏置问题，去除了HMM中两个不合理的假设。当然，模型相应得也变复杂了。

HMM

HMM模型将标注看作马尔可夫链，一阶马尔可夫链式针对相邻标注的关系进行建模，其中每个标记对应一个概率函数。HMM是一种产生式模型（generative model），定义了联合概率分布，

其中x和y分别表示观察序列和相对应的标注序列的随机变量。为了能够定义这种联合概率分布，产生式模型需要枚举出所有可能的观察序列，这在实际运算过程中很困难，因为我们需要将观察序列的元素看做是彼此孤立的个体即假设每个元素彼此独立，任何时刻的观察结果只依赖于该时刻的状态。

HMM模型的这个假设前提在比较小的数据集上是合适的，但实际上在大量真实语料中观察序列更多的是以一种多重的交互特征形式表现，观察元素之间广泛存在长程相关性。

在命名实体识别的任务中，由于实体本身结构所具有的复杂性，利用简单的特征函数往往无法涵盖所有的特性，这时HMM的假设前提使得它无法使用复杂特征(它无法使用多于一个标记的特征。)

MEMM

最大熵模型可以使用任意的复杂相关特征，在性能上最大熵分类器超过了Byaes分类器。但是，作为一种分类器模型，这两种方法有一个共同的缺点：

每个词都是单独进行分类的，标记之间的关系无法得到充分利用，具有马尔可夫链的HMM模型可以建立标记之间的马尔可夫关联性，这是最大熵模型所没有的。

最大熵模型的优点：

首先，最大熵统计模型获得的是所有满足约束条件的模型中信息熵极大的模型;

其次，最大熵统计模型可以灵活地设置约束条件，通过约束条件的多少可以调节模型对未知数据的适应度和对已知数据的拟合程度;

再次，它还能自然地解决了统计模型中参数平滑的问题。

最大熵模型的不足：

首先，最大熵统计模型中二值化特征只是记录特征的出现是否，而文本分类需要知道特征的强度，因此，它在分类方法中不是最优的;

其次，由于算法收敛的速度较慢，所以导致最大熵统计模型它的计算代价较大，时空开销大;再次，数据稀疏问题比较严重。

最大熵马尔科夫模型把HMM模型和maximum-entropy模型的优点集成为一个产生式模型，这个模型允许状态转移概率依赖于序列中彼此之间非独立的特征上，

从而将上下文信息引入到模型的学习和识别过程中，提高了识别的精确度，召回率也大大的提高，有实验证明，这个新的模型在序列标注任务上表现的比HMM和无状态的最大熵模型要好得多。

CRF

CRF模型的特点：

首先，CRF在给定了观察序列的情况下，对整个的序列的联合概率有一个统一的指数模型。一个比较吸引人的特性是其 损失函数 的凸面性。

其次，条件随机域模型相比较改进的隐马尔可夫模型可以更好更多的利用待识别文本中所提供的上下文信息以得更好的实验结果。

条件随机域在中文组块识别方面有效，并避免了严格的独立性假设和数据归纳偏置问题。

条件随机域(CRF)模型应用到了中文名实体识别中，并且根据中文的特点，定义了多种特征模板。并且有测试结果表明：在采用相同特征集合的条件下，条件随机域模型较其他概率模型有更好的性能表现。

再次，词性标注主要面临兼类词消歧以及未知词标注的难题，传统隐马尔科夫方法不易融合新特征，而最大熵马尔科夫模型存在标注偏置等问题。

论文引入条件随机域建立词性标注模型，易于融合新的特征，并能解决标注偏置的问题。

CRFs具有很强的推理能力，并且能够使用复杂、有重叠性和非独立的特征进行训练和推理，能够充分地利用上下文信息作为特征，还可以任意地添加其他外部特征，

使得模型能够获取的信息非常丰富。同时，CRFs解决了最大熵模型中的“label bias”问题。

CRFs与最大熵模型的本质区别是：最大熵模型在每个状态都有一个概率模型，在每个状态转移时都要进行归一化。如果某个状态只有一个后续状态，那么该状态到后续状态的跳转概率即为1。这样，不管输入为任何内容，它都向该后续状态跳转。而CRFs是在所有的状态上建立一个统一的概率模型，这样在进行归一化时，即使某个状态只有一个后续状态，它到该后续状态的跳转概率也不会为1，从而解决了“label bias”问题。因此，从理论上讲，CRFs非常适用于中文的词性标注。

CRF模型的优点：

首先，CRF模型由于其自身在结合多种特征方面的优势和避免了标记偏置问题。

其次，CRF的性能更好，CRF对特征的融合能力比较强，

对于实例较小的时间类ME来说，CRF的识别效果明显高于ME的识别结果。

CRF模型的不足：首先，通过对基于CRF的结合多种特征的方法识别英语命名实体的分析，发现在使用CRF方法的过程中，特征的选择和优化是影响结果的关键因素，

特征选择问题的好坏，直接决定了系统性能的高低。其次，训练模型的时间比ME更长，且获得的模型很大，在一般的PC机上无法运行。