

## 条件随机场CRF(一)从随机场到线性链条件随机场

[条件随机场CRF\(一\)从随机场到线性链条件随机场](#)

[条件随机场CRF\(二\) 前向后向算法评估标记序列概率](#)

[条件随机场CRF\(三\) 模型学习与维特比算法解码](#)

条件随机场(Conditional Random Fields, 以下简称CRF)是给定一组输入序列条件下另一组输出序列的条件概率分布模型，在自然语言处理中得到了广泛应用。本系列主要关注于CRF的特殊形式：线性链(Linear chain) CRF。本文关注与CRF的模型基础。

# 1.什么样的问题需要CRF模型

和HMM类似，在讨论CRF之前，我们来看看什么样的问题需要CRF模型。这里举一个简单的例子：

假设我们有Bob一天从早到晚的一系列照片，Bob想考考我们，要我们猜这一系列的每张照片对应的活动，比如：工作的照片，吃饭的照片，唱歌的照片等等。一个比较直观的办法就是，我们找到Bob之前的日常生活的一系列照片，然后找Bob问清楚这些照片代表的活动标记，这样我们就可以用监督学习的方法来训练一个分类模型，比如逻辑回归，接着用模型去预测这一天的每张照片最可能的活动标记。

这种办法虽然是可行的，但是却忽略了一个重要的问题，就是这些照片之间的顺序其实是有很大的时间顺序关系的，而用上面的方法则会忽略这种关系。比如我们现在看到了一张Bob闭着嘴的照片，那么这张照片我们怎么标记Bob的活动呢？比较难去打标记。但是如果我们Bob在这一张照片前一点点时间的照片的话，那么这张照片就好标记了。如果在时间序列上前一一张的照片里Bob在吃饭，那么这张闭嘴的照片很有可能是在吃饭咀嚼。而如果在时间序列上前一一张的照片里Bob在唱歌，那么这张闭嘴的照片很有可能是在唱歌。

### 公告

★珠江追梦，饮岭南茶，恋鄂北家★

昵称：刘建平Pinard

园龄：1年7个月

粉丝：1502

关注：14

[+加关注](#)

### 随笔分类(105)

0040. 数学统计学(4)

0081. 机器学习(66)

0082. 深度学习(10)

0083. 自然语言处理(23)

0121. 大数据挖掘(1)

0122. 大数据平台(1)

0123. 大数据可视化

### 随笔档案(105)

为了让我们的分类器表现的更好，可以在标记数据的时候，可以考虑相邻数据的标记信息。这一点，是普通的分类器难以做到的。而这一块，也是CRF比较擅长的地方。

在实际应用中，自然语言处理中的词性标注(POS Tagging)就是非常适合CRF使用的地方。词性标注的目标是给出一个句子中每个词的词性（名词，动词，形容词等）。而这些词的词性往往和上下文的词的词性有关，因此，使用CRF来处理是很适合的，当然CRF不是唯一的选择，也有很多其他的词性标注方法。

## 2. 从随机场到马尔科夫随机场

首先，我们来看看什么是随机场。“随机场”的名字取的很玄乎，其实理解起来不难。随机场是由若干个位置组成的整体，当给每一个位置中按照某种分布随机赋予一个值之后，其全体就叫做随机场。还是举词性标注的例子：假如我们有一个十个词形成的句子需要做词性标注。这十个词每个词的词性可以在我们已知的词性集合（名词，动词...）中去选择。当我们为每个词选择完词性后，这就形成了一个随机场。

了解了随机场，我们再来看看马尔科夫随机场。马尔科夫随机场是随机场的特例，它假设随机场中某一个位置的赋值仅仅与和它相邻的位置的赋值有关，和与其不相邻的位置的赋值无关。继续举十个词的句子词性标注的例子：如果我们假设所有词的词性只和它相邻的词的词性有关时，这个随机场就特化成一个马尔科夫随机场。比如第三个词的词性除了与自己本身的位置有关外，只与第二个词和第四个词的词性有关。

## 3. 从马尔科夫随机场到条件随机场

理解了马尔科夫随机场，再理解CRF就容易了。CRF是马尔科夫随机场的特例，它假设马尔科夫随机场中只有 $X$ 和 $Y$ 两种变量， $X$ 一般是给定的，而 $Y$ 一般是在给定 $X$ 的条件下我们的输出。这样马尔科夫随机场就特化成了条件随机场。在我们十个词的句子词性标注的例子中， $X$ 是词， $Y$ 是词性。因此，如果我们假设它是一个马尔科夫随机场，那么它也就是一个CRF。

对于CRF，我们给出准确的数学语言描述：

设 $X$ 与 $Y$ 是随机变量， $P(Y|X)$ 是给定 $X$ 时 $Y$ 的条件概率分布，若随机变量 $Y$ 构成的是一个马尔科夫随机场，则称条件概率分布 $P(Y|X)$ 是条件随机场。

## 4. 从条件随机场到线性链条件随机场

注意在CRF的定义中，我们并没有要求 $X$ 和 $Y$ 有相同的结构。而实现中，我们一般都假设 $X$ 和 $Y$ 有相同的结构，即：

$$X = (X_1, X_2, \dots, X_n), Y = (Y_1, Y_2, \dots, Y_n)$$

2018年6月 (1)

2018年5月 (3)

2017年8月 (1)

2017年7月 (3)

2017年6月 (8)

2017年5月 (7)

2017年4月 (5)

2017年3月 (10)

2017年2月 (7)

2017年1月 (13)

2016年12月 (17)

2016年11月 (22)

2016年10月 (8)

### 常去的机器学习网站

52 NLP  
Analytics Vidhya  
机器学习库  
机器学习路线图  
深度学习进阶书  
深度学习入门书

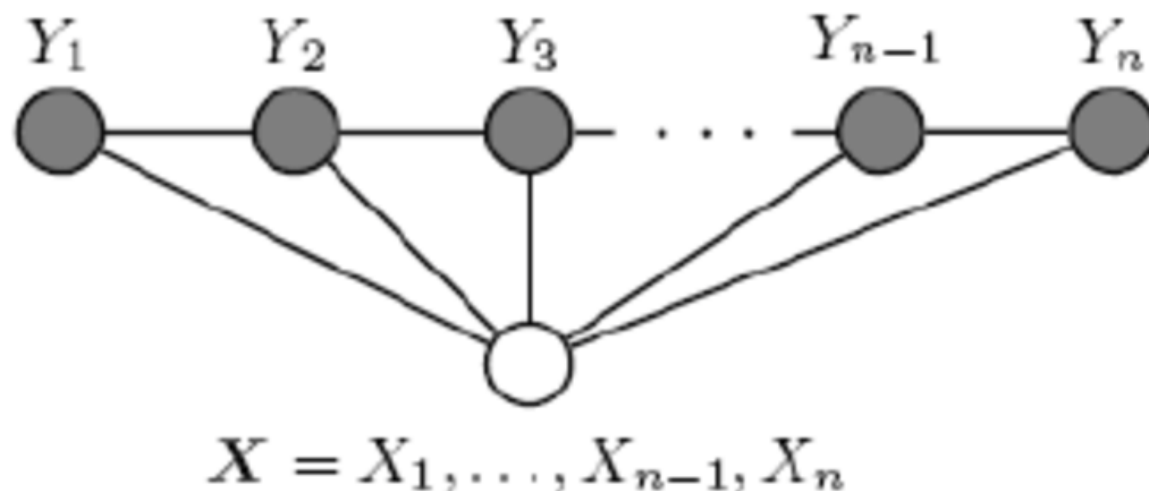
### 积分与排名

积分 - 314755  
排名 - 595

### 阅读排行榜

1. 梯度下降（Gradient Descent）小结(131942)
2. 梯度提升树(GBDT)原理小结(72360)
3. 线性判别分析LDA原理总结(48861)

我们一般考虑如下图所示的结构： $X$ 和 $Y$ 有相同的结构的CRF就构成了线性链条件随机场(Linear chain Conditional Random Fields,以下简称 linear-CRF)。



在我们的十个词的句子的词性标记中，词有十个，词性也是十个，因此，如果我们假设它是一个马尔科夫随机场，那么它也就是一个linear-CRF。

我们再来看看 linear-CRF的数学定义：

设 $X = (X_1, X_2, \dots, X_n)$ ， $Y = (Y_1, Y_2, \dots, Y_n)$ 均为线性链表示的随机变量序列，在给定随机变量序列 $X$ 的情况下，随机变量 $Y$ 的条件概率分布 $P(Y|X)$ 构成条件随机场，即满足马尔科夫性：

$$P(Y_i|X, Y_1, Y_2, \dots, Y_n) = P(Y_i|X, Y_{i-1}, Y_{i+1})$$

则称 $P(Y|X)$ 为线性链条件随机场。

## 5. 线性链条件随机场的参数化形式

对于上一节讲到的linear-CRF，我们如何将其转化为可以学习的机器学习模型呢？这是通过特征函数和其权重系数来定义的。什么是特征函数呢？

在linear-CRF中，特征函数分为两类，第一类是定义在 $Y$ 节点上的节点特征函数，这类特征函数只和当前节点有关，记为：

4. scikit-learn决策树算法类库使用小结(37338)
5. 循环神经网络(RNN)模型与前向反向传播算法(34727)

### 评论排行榜

1. 梯度提升树(GBDT)原理小结(135)
2. 谱聚类 ( spectral clustering ) 原理总结(93)
3. 梯度下降 ( Gradient Descent ) 小结(85)
4. 集成学习之Adaboost算法原理小结(82)
5. 卷积神经网络(CNN)反向传播算法(65)

### 推荐排行榜

1. 梯度下降 ( Gradient Descent ) 小结(48)
2. 奇异值分解(SVD)原理与在降维中的应用(23)
3. 集成学习之Adaboost算法原理小结(16)
4. 集成学习原理小结(16)
5. 卷积神经网络(CNN)反向传播算法(16)

$$s_l(y_i, x, i), \quad l = 1, 2, \dots, L$$

其中 $L$ 是定义在该节点的节点特征函数的总个数， $i$ 是当前节点在序列的位置。

第二类是定义在 $Y$ 上下文的局部特征函数，这类特征函数只和当前节点和上一个节点有关，记为：

$$t_k(y_{i-1}, y_i, x, i), \quad k = 1, 2, \dots, K$$

其中 $K$ 是定义在该节点的局部特征函数的总个数， $i$ 是当前节点在序列的位置。之所以只有上下文相关的局部特征函数，没有不相邻节点之间的特征函数，是因为我们的linear-CRF满足马尔科夫性。

无论是节点特征函数还是局部特征函数，它们的取值只能是0或者1。即满足特征条件或者不满足特征条件。同时，我们可以为每个特征函数赋予一个权值，用以表达我们对这个特征函数的信任度。假设 $t_k$ 的权重系数是 $\lambda_k$ ， $s_l$ 的权重系数是 $\mu_l$ ，则linear-CRF由我们所有的 $t_k, \lambda_k, s_l, \mu_l$ 共同决定。

此时我们得到了linear-CRF的参数化形式如下：

$$P(y|x) = \frac{1}{Z(x)} \exp\left(\sum_{i,k} \lambda_k t_k(y_{i-1}, y_i, x, i) + \sum_{i,l} \mu_l s_l(y_i, x, i)\right)$$

其中， $Z(x)$ 为规范化因子：

$$Z(x) = \sum_y \exp\left(\sum_{i,k} \lambda_k t_k(y_{i-1}, y_i, x, i) + \sum_{i,l} \mu_l s_l(y_i, x, i)\right)$$

回到特征函数本身，每个特征函数定义了一个linear-CRF的规则，则其系数定义了这个规则的可信度。所有的规则和其可信度一起构成了我们的linear-CRF的最终的条件概率分布。

## 6. 线性链条件随机场实例

这里我们给出一个linear-CRF用于词性标注的实例，为了方便，我们简化了词性的种类。假设输入的都是三个词的句子，即 $X = (X_1, X_2, X_3)$ ，输出的词性标记为 $Y = (Y_1, Y_2, Y_3)$ ，其中 $Y \in \{1(\text{名词}), 2(\text{动词})\}$

这里只标记出取值为1的特征函数如下：

$$t_1 = t_1(y_{i-1} = 1, y_i = 2, x, i), i = 2, 3, \quad \lambda_1 = 1$$

$$t_2 = t_2(y_1 = 1, y_2 = 1, x, 2) \quad \lambda_2 = 0.5$$

$$t_3 = t_3(y_2 = 2, y_3 = 1, x, 3) \quad \lambda_3 = 1$$

$$t_4 = t_4(y_1 = 2, y_2 = 1, x, 2) \quad \lambda_4 = 1$$

$$t_5 = t_5(y_2 = 2, y_3 = 2, x, 3) \quad \lambda_5 = 0.2$$

$$s_1 = s_1(y_1 = 1, x, 1) \quad \mu_1 = 1$$

$$s_2 = s_2(y_i = 2, x, i), i = 1, 2, \quad \mu_2 = 0.5$$

$$s_3 = s_3(y_i = 1, x, i), i = 2, 3, \quad \mu_3 = 0.8$$

$$s_4 = s_4(y_3 = 2, x, 3) \quad \mu_4 = 0.5$$

求标记(1,2,2)的非规范化概率。

利用linear-CRF的参数化公式，我们有：

$$P(y|x) \propto \exp \left[ \sum_{k=1}^5 \lambda_k \sum_{i=2}^3 t_k(y_{i-1}, y_i, x, i) + \sum_{l=1}^4 \mu_l \sum_{i=1}^3 s_l(y_i, x, i) \right]$$

带入(1,2,2)我们有：

$$P(y_1 = 1, y_2 = 2, y_3 = 2|x) \propto \exp(3.2)$$

## 7. 线性链条件随机场的简化形式

在上几节里面，我们用 $s_l$ 表示节点特征函数，用 $t_k$ 表示局部特征函数，同时也用了不同的符号表示权重系数，导致表示起来比较麻烦。其实我们可以对特征函数稍加整理，将其统一起来。

假设我们在某一节点我们有 $K_1$ 个局部特征函数和 $K_2$ 个节点特征函数，总共有 $K = K_1 + K_2$ 个特征函数。我们用一个个特征函数 $f_k(y_{i-1}, y_i, x, i)$ 来统一表示如下：

$$f_k(y_{i-1}, y_i, x, i) = \begin{cases} t_k(y_{i-1}, y_i, x, i) & k = 1, 2, \dots, K_1 \\ s_l(y_i, x, i) & k = K_1 + l, l = 1, 2, \dots, K_2 \end{cases}$$

对 $f_k(y_{i-1}, y_i, x, i)$ 在各个序列位置求和得到：

$$f_k(y, x) = \sum_{i=1}^n f_k(y_{i-1}, y_i, x, i)$$

同时我们也统一 $f_k(y_{i-1}, y_i, x, i)$ 对应的权重系数 $w_k$ 如下：

$$w_k = \begin{cases} \lambda_k & k = 1, 2, \dots, K_1 \\ \mu_l & k = K_1 + l, l = 1, 2, \dots, K_2 \end{cases}$$

这样，我们的linear-CRF的参数化形式简化为：

$$P(y|x) = \frac{1}{Z(x)} \exp \sum_{k=1}^K w_k f_k(y, x)$$

其中， $Z(x)$ 为规范化因子：

$$Z(x) = \sum_y \exp \sum_{k=1}^K w_k f_k(y, x)$$

如果将上两式中的 $w_k$ 与 $f_k$ 的用向量表示，即：

$$w = (w_1, w_2, \dots, w_K)^T \quad F(y, x) = (f_1(y, x), f_2(y, x), \dots, f_K(y, x))^T$$

则linear-CRF的参数化形式简化为内积形式如下：

$$P_w(y|x) = \frac{\exp(w \bullet F(y, x))}{Z_w(x)} = \frac{\exp(w \bullet F(y, x))}{\sum_y \exp(w \bullet F(y, x))}$$

## 8. 线性链条件随机场的矩阵形式

将上一节统一后的linear-CRF公式加以整理，我们还可以将linear-CRF的参数化形式写成矩阵形式。为此我们定义一个  $m \times m$  的矩阵  $M$ ， $m$  为  $y$  所有可能的状态的取值个数。 $M$  定义如下：

$$M_i(x) = [M_i(y_{i-1}, y_i | x)] = [\exp(W_i(y_{i-1}, y_i | x))] = \left[ \exp \left( \sum_{k=1}^K w_k f_k(y_{i-1}, y_i, x, i) \right) \right]$$

我们引入起点和终点标记  $y_0 = start, y_{n+1} = stop$ ，这样，标记序列  $y$  的非规范化概率可以通过  $n + 1$  个矩阵元素的乘积得到，即：

$$P_w(y|x) = \frac{1}{Z_w(x)} \prod_{i=1}^{n+1} M_i(y_{i-1}, y_i | x)$$

其中  $Z_w(x)$  为规范化因子。

以上就是linear-CRF的模型基础，后面我们会讨论linear-CRF和HMM类似的三个问题的求解方法。

( 欢迎转载，转载请注明出处。欢迎沟通交流：liujianping-ok@163.com )

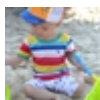
分类: [0083. 自然语言处理](#)

标签: [自然语言处理](#)

好文要顶

关注我

收藏该文



[刘建平Pinard](#)

[关注 - 14](#)

[粉丝 - 1502](#)

[+加关注](#)

4

0

« 上一篇: [用hmmlearn学习隐马尔科夫模型HMM](#)

» 下一篇: [条件随机场CRF\(二\) 前向后向算法评估标记序列概率](#)

---

posted @ 2017-06-19 17:32 刘建平Pinard 阅读(5951) 评论(11) 编辑 收藏

## 评论列表

---

#1楼 2017-10-24 21:25 jackyxu\_1111

X, Y具有相同的结构, 这个不一定。

支持(0) 反对(0)

---

#2楼[楼主] 2017-10-26 10:28 刘建平Pinard

@ jackyxu\_1111

你好, 的确不要求相同的结构, 只不过我上文中只讨论了X和Y结构相同时候的情况。

支持(0) 反对(0)

---

#3楼 2017-12-21 10:06 / 小黄豆

请问在第六条线性链条件随机场实例中，您说了这里只标记出取值为1的特征函数，但为什么在您列出的公式中有的当前序列的y的结果为2而不是仅仅只有1？

支持(0) 反对(0)

---

#4楼[楼主] 2017-12-21 11:11 刘建平Pinard

@ 小黄豆

你好，这里取值为1的意思是这个特征函数在某些序列的位置成立，即针对函数 $t, s$ 的值。如果这个特征函数在某些序列的位置不成立，那么 $t, s$ 取值为0.

还是以这一节的例子中的 $t_1$ 函数为例，我们里面有这样的式子：

$$t_1 = t_1(y_{i-1} = 1, y_i = 2, x, i), i = 2, 3, \lambda_1 = 1$$

即在 $y_{i-1} = 1, y_i = 2, i = 2, 3$ 这个条件下 $t_1 = 1$ 。

对于 $t_1$ 函数，在 $i=1$ 的时候，由于不成立，那么此时 $t_1 = 0$ ，这样的特征函数我们没有标出。

支持(2) 反对(0)

---

#5楼 2018-03-01 13:15 littlely

你好，能否解释一下 $M_i$ 矩阵意思， $\mathbf{y}$ 内是矩阵中的每一个元素吗？又怎么解释其中的各个元素？谢谢

支持(0) 反对(0)

---

#6楼[楼主] 2018-03-02 10:56 刘建平Pinard

@ littlely

你好，是的， $\mathbf{y}$ 是矩阵中的每个元素。

$M_i$ 矩阵中的每个元素对应于 $\mathbf{y}$ 中的 $\exp(\sum_{k=1}^K w_k f_k(y_{i-1}, y_i, x, i))$ ，即对于 $y$ 中每个取值 $y_i$ ，可以得到一个非规范化概率的矩阵，矩阵中的每一个点对应从任意一个 $y$ 取值到取值 $y_i$ 的非规范化概率。

支持(0) 反对(0)

---

#7楼 2018-03-10 20:53 一夜听雨风萧萧

您好，看完之后收益匪浅，但是要应用到具体的实例上，所以请问有实现CRF的代码吗？python或者java？

支持(0) 反对(0)

---

#8楼[楼主] 2018-03-12 10:35 刘建平Pinard



@ 一夜听雨风萧萧

你好，如果你要用在实例上，推荐使用CRFSuite.

<http://www.chokkan.org/software/crfsuite/>

支持(0) 反对(0)

---

#9楼 2018-03-12 11:47 学习并长成

博主您好，我有个疑问，就是在 $Z(X)$ 的计算公式上。

在第5节中说到 $Z(X)$ 是归一因子，对 $y$ 求和，那么我觉得既然已经对 $y$ 求和了，里面的 $\exp(\cdot)$ 部分中，就不应该再对 $i$ 求和了。不知道我的理解是否有问题。谢谢。

支持(0) 反对(0)

---

#10楼[楼主] 2018-03-12 14:42 刘建平Pinard

@ 学习并长成

你好，分子里代表的单是 $P(Y_i|X, Y_{i-1})$ 的未归一化概率，即单表示了 $Y_i$ 某一种取值的概率分布，而分母的归一化因子可以理解为对 $Y$ 的所有状态 $Y_1, Y_2, \dots, Y_i, \dots, Y_N$ 的未归一化概率分布求和。

分子的求和部分针对的是 $Y_i$ 的特征函数，并不针对所有的状态序号 $i$ 取值。主要是特征函数是和具体的状态序号 $i$ 绑定的，所以可以看到求和里有 $i$ ，但不意味着要对所有的状态序号 $i$ 对应的值求和。

支持(0) 反对(0)

---

#11楼 2018-04-06 00:15 jasenchnn

@ 学习并长成

感觉这里的问题可以参考这篇帖子：

条件随机场——时间序列（句子单词序列也算），其特征函数必须要考虑前一刻的数据

没法添加链接，建议搜索下，从知乎摘的

bty：楼主的帖子讲的真好，谢谢楼主

支持(0) 反对(0)

---

[刷新评论](#) [刷新页面](#) [返回顶部](#)

注册用户登录后才能发表评论，请 [登录](#) 或 [注册](#)，[访问网站首页](#)。