

CRF-条件随机场

原创

2016年11月14日 07:56:29

标签

机器学习 / 算法 / NLP / 动态规划

436

注：本文中所有公式和思路来自于邹博先生的《机器学习升级版》，我只是为了加深记忆和理解写的本文。

这次要分享的是模型是条件随机场模型，听起来就是不是十分的容易，不过也确实是这样子的，学习条件随机场之前个人觉得最好先学习HMM模型，因为条件随机场和HMM有很多共同的地方，比方说都是概率图模型、最基本的三个问题也是一模一样的(概率计算、学习问题、预测问题)。

这里先给出一个条件随机场的定义：条件随机场时给定一组输入随机变量X的条件下另一组随机变量Y的条件概率分布模型，其特点是假设输出随机变量Y构成马尔科夫随机场。大家一定记得Logistic吧，Logistic这个模型中不也是给定输入随机变量X的条件下另一组随机变量Y的条件概率么，只不过Logistic的输出随机变量时离散的、相互之间是独立的，而CRF(条件随机场)的输出随机变量之间我们是假设构成马尔科夫随机场的，仅此而已，所以我们可以套用Logistic的套路来搞点事情。

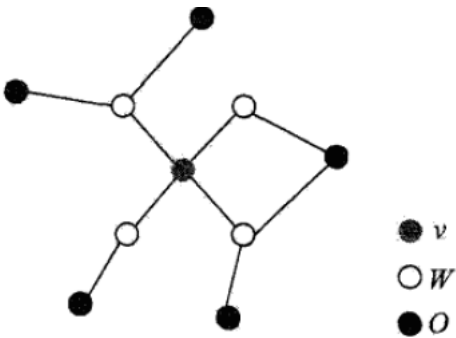
那么什么是马尔科夫随机场呢？

概率无向图模型：设有联合概率分布 $p(Y)$ ，由无向图 $G=(V,E)$ 表示，在图 G 中， V (结点)表示随机变量， E (边)表示随机变量之间的依赖关系，如果联合概率 $P(Y)$ 满足成对、局部、全局马尔科夫性，就称此联合概率分布为概率无向图模型(probability undirected graphical model)，或马尔科夫随机场(Markov random filed)

那么什么是成对、局部、全局马尔科夫性呢？

我只说一下局部和全局马尔科夫性，成对的极其简单，和局部、全局都是一个道理。

局部马尔科夫性：



稍作解释：图中分为三部分v、w、o，如果给定w，那么v和o就是独立的，公式化的解释是：



重庆旅游攻略



联系我们



请扫描二维码联系
webmaster@
400-660-011
QQ客服

关于 招聘 广告服务
©1999-2018 CSDN版权所有
京ICP证09002463号

经营性网站备案信息
网络110报警服务
中国互联网举报中心
北京互联网违法和不良信息举报中心

他的最新文章

- 压力测试工具siege的用法
- 【卷积神经网络-进化史】从LeNet
- linux 环境变量设置（临时 + 永
- gtest install
- linux(ubuntu)下的caffe编译安

文章分类

- 机器学习
- linux
- deep learning

文章存档

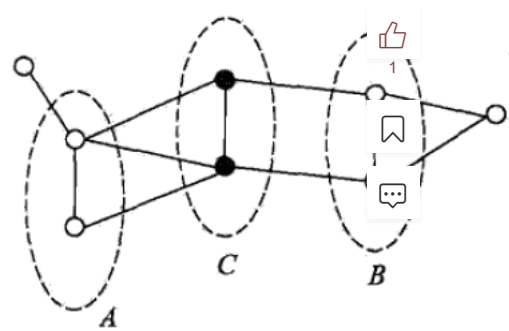
- 2017年1月
- 2016年12月
- 2016年11月

他的热门文章

- linux(ubuntu)下的caffe编译安
1748
- EM算法---基于隐变量的参数估计
1674
- 概率图模型之贝叶斯网络
1657

$$P(Y_v, Y_o | Y_w) = P(Y_v | Y_w)P(Y_o | Y_w)$$

全局马尔科夫性：



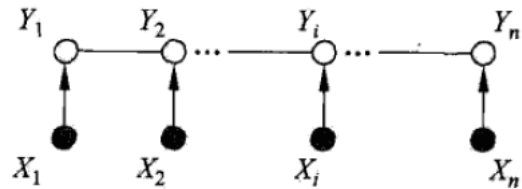
解释：给定三个集合A、B、C，如果给定C集合，那么A、B两个集合就是条件独立的，公式化解释：

$$P(Y_A, Y_B | Y_C) = P(Y_A | Y_C)P(Y_B | Y_C)$$

成对、局部、全局马尔科夫性是等价的，后两个是成对马尔科夫性推广而来的。此处文献可以查阅李航博士的《统计学习方法》。

条件随机场有多种形式，本文只介绍线性链条件随机场，也就是说输出随机变量是在一条链上的，我们可以这么来理解。

下面这种是我们最常见的一种线性链条件随机场



从Logistic谈起

假定：

$$\begin{aligned} P(y = 1 | x; \theta) &= h_{\theta}(x) \\ P(y = 0 | x; \theta) &= 1 - h_{\theta}(x) \end{aligned}$$

则有：



重庆旅游攻略



联系我们



请扫描二维码联系

✉ webmaster@

☎ 400-660-010

🗣 QQ客服

关于 招聘 广告服务

©1999-2018 CSDN版权所有

京ICP证09002463号

经营性网站备案信息

网络110报警服务

中国互联网举报中心

北京互联网违法和不良信息举报中心

$$p(y | x; \theta) = (h_{\theta}(x))^y (1 - h_{\theta}(x))^{1-y}$$

$$\begin{aligned} L(\theta) &= p(\vec{y} | X; \theta) \\ &= \prod_{i=1}^m p(y^{(i)} | x^{(i)}; \theta) \\ &= \prod_{i=1}^m (h_{\theta}(x^{(i)})^{y^{(i)}} (1 - h_{\theta}(x^{(i)}))^{1-y^{(i)}}) \end{aligned}$$

对数似然：

$$\begin{aligned} \ell(\theta) &= \log L(\theta) \\ &= \sum_{i=1}^m y^{(i)} \log h(x^{(i)}) + (1 - y^{(i)}) \log(1 - h(x^{(i)})) \end{aligned}$$

对参数求导：

$$\begin{aligned} \frac{\partial}{\partial \theta_j} \ell(\theta) &= \left(y \frac{1}{g(\theta^T x)} - (1 - y) \frac{1}{1 - g(\theta^T x)} \right) \frac{\partial}{\partial \theta_j} g(\theta^T x) \\ &= \left(y \frac{1}{g(\theta^T x)} - (1 - y) \frac{1}{1 - g(\theta^T x)} \right) g(\theta^T x) (1 - g(\theta^T x)) \frac{\partial}{\partial \theta_j} \theta^T x \\ &= (y(1 - g(\theta^T x)) - (1 - y)g(\theta^T x)) x_j \\ &= (y - h_{\theta}(x)) x_j \end{aligned}$$

Logistic回归的参数学习过程：

$$\theta_j := \theta_j + \alpha (y^{(i)} - h_{\theta}(x^{(i)})) x_j^{(i)}$$

我们都知道一件事情发生的几率是该事件发生的概率与不发生概率的比值，那么对数几率就可以这么写：

$$\text{logit}(p) = \log \frac{p}{1-p} = \log \frac{h_{\theta}(x)}{1-h_{\theta}(x)} = \log \left(\frac{1}{1+e^{-\theta^T x}} \frac{e^{-\theta^T x}}{1+e^{-\theta^T x}} \right) = \theta^T x$$

那么对数线性模型的一般形式可以这么写：

$$p(y | x; w) = \frac{1}{Z(x, w)} \exp \left(\sum_j w_j F_j(x, y) \right)$$

其中我们将Logistic/Softmax回归的特征记作 $F_j(x, y)$ ，归一化因子为：

加入CSDN，享受更精准的内容推荐，与500万程序员共同成长！



重庆旅游攻略



联系我们



请扫描二维码联系

webmaster@csdn.net

400-660-0111

QQ客服 95511

关于 招聘 广告服务 网站地图

©1999-2018 CSDN版权所有

京ICP证09002463号

经营性网站备案信息

网络110报警服务

中国互联网举报中心

北京互联网违法和不良信息举报中心

$$Z(x, w) = \sum_y \exp \sum_j w_j F_j(x, y)$$

给定一个x的预测标记为：

$$\hat{y} = \arg \max_y p(y | x, w) = \max_y \sum_j w_j F_j(x, y)$$

当然了，CRF毕竟跟Logistic有很大的不同，但思路我们不妨借鉴一下Logistic。

线性链条件随机场可以使用对数线性模型来做，使用 \bar{x} 来表示n个词的序列，使用 \bar{y} 表示相应的词性：

$$p(\bar{y} | \bar{x}; w) = \frac{1}{Z(\bar{x}, w)} \exp \left(\sum_j w_j F_j(\bar{x}, \bar{y}) \right)$$

毕竟和Logistic不同，我们定义 \bar{x} 的第j个特征 F_j 是有若干的次特征组成：

$$f_j(y_{i-1}, y_i, \bar{x}, i)$$

这里 f_j 依赖或者部分依赖于当前的整个句子 \bar{x} 、当前的词的标记 y_i 、前一个词的标记 y_{i-1} ，当前词在句子中的位置 i ：

$$F_j(\bar{x}, \bar{y}) = \sum_i f_j(y_{i-1}, y_i, \bar{x}, i)$$

那么我们接下来就要找出参数，满足下式的 w ，即为最终所求参数：

$$\bar{y}^* = \arg \max_{\bar{y}} p(\bar{y} | \bar{x}, w)$$

我们进一步将前边的公式带入：

$$\bar{y}^* = \arg \max_{\bar{y}} p(\bar{y} | \bar{x}, w) = \arg \max_{\bar{y}} \sum_j w_j F_j(\bar{x}, \bar{y})$$

这里是省略了归一化因子和指数e，因为对我们求最值来说没影响，所以简化了一下，进一步代入公式：

$$\begin{aligned} \bar{y}^* &= \arg \max_{\bar{y}} \sum_j w_j \sum_i f_j(y_{i-1}, y_i, \bar{x}, i) = \arg \max_{\bar{y}} \sum_j \sum_i w_j f_j(y_{i-1}, y_i, \bar{x}, i) \\ &= \arg \max_{\bar{y}} \sum_i \sum_j w_j f_j(y_{i-1}, y_i, \bar{x}, i) \\ &= \arg \max_{\bar{y}} \sum_i g_j(y_{i-1}, y_i) \quad g_j(y_{i-1}, y_i) = \sum_j w_j f_j(y_{i-1}, y_i, \bar{x}, i) \end{aligned}$$



重庆旅游攻略



联系我们



请扫描二维码联系

webmaster@

400-660-011

QQ客服

关于 招聘 广告服务

©1999-2018 CSDN版权所有

京ICP证09002463号

经营性网站备案信息

网络110报警服务

中国互联网举报中心

北京互联网违法和不良信息举报中心

其中第一行第二个等号只不过是加和w换了位置，不影响结果，第二行是将两个加和互换位置，也不会对结果有影响，最后一行是做了一个定义，也就是g函数，其实这块有一处有问题，应该是gi而不是gj，因为关于i的一个函数，其实无关紧要，接着往下。

接着我们令 $\alpha_k(v)$ 为前向得分，表示第k个词的状态v的最大的分(归一化后为概率)：

$$\alpha_k(v) = \max_{y_1, y_2, \dots, y_{k-1}} \left(\sum_{i=1}^{k-1} g_i(y_{i-1}, y_i) + g_k(y_{k-1}, v) \right)$$

这块其实我们在别的资料上被介绍为t和s函数，其实是一样的，gi或者说s是关于边的转移特征函数，gk或者t是关于结点的状态特征函数，那么我们的递推公式为：

$$\alpha_k(v) = \max_{y_{k-1}} (\alpha_{k-1}(y_{k-1}) + g_k(y_{k-1}, v))$$

时间复杂度是 $O(m^2n)$ ，其中标记数为m，句子中包含的单词数目为n。

那该怎么计算概率呢？

由于：

$$g_j(y_{i-1}, y_i) = \sum_j w_j f_j(y_{i-1}, y_i, \bar{x}, i)$$

得：

$$\begin{aligned} Z(\bar{x}, w) &= \sum_{\bar{y}} \exp \sum_j w_j F_j(\bar{x}, \bar{y}) \\ &= \sum_{\bar{y}} \exp \sum_i g_j(y_{i-1}, y_i) \\ &= \sum_{\bar{y}} \prod_i \exp(g_j(y_{i-1}, y_i)) \end{aligned}$$

从最后一步公式中，我们可以看出，实现对一堆东西算乘积再加和，这不就矩阵乘法嘛，所以我们定义一个m*m得矩阵：

$$M_i(u, v) = \exp(g_i(u, v))$$

那么对于M1，我们任选一个u=start，Mn+1任选一个v=stop，那么从M1到M2可以如下表示：



加入CSDN，享受更精准的内容推荐，与500万程序员共同成长！

登录

注册



重庆旅游攻略



联系我们



请扫描二维码联系

webmaster@

400-660-011

QQ客服

关于 招聘 广告服务

©1999-2018 CSDN版权所有

京ICP证09002463号

经营性网站备案信息

网络110报警服务

中国互联网举报中心

北京互联网违法和不良信息举报中心

$$\begin{aligned}
 & M_{12}(start, v) \\
 &= \sum_q M_1(start, q) M_2(q, v) \\
 &= \sum_q \exp(g_1(start, q)) \exp(g_2(q, v))
 \end{aligned}$$



推广到n+1：

$$\begin{aligned}
 M_{1,2,3,\dots,n+1}(start, stop) &= \sum_{y_1, y_2, \dots, y_n} M_1(start, y_1) M_2(y_1, y_2) \cdots M_{n+1}(y_n, stop) \\
 &= \sum_{\bar{y}} \prod_i \exp(g_j(y_{i-1}, y_i))
 \end{aligned}$$

那么现在回到之前的求参数得问题上，我们是要求出权向量 w ，使得下面得式子成立：

$$\bar{y}^* = \arg \max_{\bar{y}} p(\bar{y} | \bar{x}, w)$$

我们可以对其取对数，然后求驻点：

$$\begin{aligned}
 p(y | x; w) &= \frac{1}{Z(x, w)} \exp\left(\sum_j w_j F_j(x, y)\right) & Z(x, w) &= \sum_{\bar{y}} \exp\left(\sum_j w_j F_j(x, y)\right) \\
 \Rightarrow \log p(y | x; w) &= \log \frac{1}{Z(x, w)} + \log \exp\left(\sum_j w_j F_j(x, y)\right) \\
 &= -\log Z(x, w) + \sum_j w_j F_j(x, y)
 \end{aligned}$$

计算梯度：

$$\begin{aligned}
 \frac{\partial}{\partial w_j} \log p(y | x; w) &= F_j(x, y) - \frac{\partial}{\partial w_j} \log Z(x, w) \\
 &= F_j(x, y) - \frac{1}{Z(x, w)} \sum_{y'} \frac{\partial}{\partial w_j} \exp\left(\sum_{j'} w_{j'} F_{j'}(x, y')\right) \\
 &= F_j(x, y) - \frac{1}{Z(x, w)} \sum_{y'} [\exp\left(\sum_{j'} w_{j'} F_{j'}(x, y')\right)] F_j(x, y') \\
 &= F_j(x, y) - \sum_{y'} F_j(x, y') \frac{\exp\left(\sum_{j'} w_{j'} F_{j'}(x, y')\right)}{\sum_{y''} \exp\left(\sum_{j''} w_{j''} F_{j''}(x, y'')\right)} \\
 &= F_j(x, y) - \sum_{y'} F_j(x, y') p(y' | x; w) \\
 &= F_j(x, y) - E_{y' \sim p(y' | x; w)}[F_j(x, y')]
 \end{aligned}$$

这个求得过程我就不一步步细讲了，其中 $\frac{\partial}{\partial w_j}$ 和 $\frac{\partial}{\partial w_{j'}}$ 只是标记而已，不是二阶导数，那么我们就可得到梯度上升得公式：



重庆旅游攻略



联系我们



请扫描二维码联系

webmaster@csdn.net

400-660-0101

QQ客服 100370800

关于 招聘 广告服务 网站地图

©1999-2018 CSDN版权所有

京ICP证09002463号

经营性网站备案信息

网络110报警服务

中国互联网举报中心

北京互联网违法和不良信息举报中心

$$w_j := w_j + \alpha(F_j(x, y) - E_{y' \sim p(y'|x;w)}[F_j(x, y')])$$

其中 α 是学习率。



1

至此CRF中的概率计算问题就完全说完了，当然，另外两个问题就是学习问题和预测问题，这两个问题请看我得HMM几篇文章中得几个算法，都是可以解决的。



欢迎批评指正！！



重庆旅游攻略



联系我们



请扫描二维码联系

webmaster@

400-660-010

QQ客服

关于 招聘 广告服务

©1999-2018 CSDN版权所有
京ICP证09002463号

经营性网站备案信息
网络110报警服务
中国互联网举报中心
北京互联网违法和不良信息举报中心

版权声明：本文为博主原创文章，未经博主允许不得转载。 <https://blog.csdn.net/u012771351/article/details/53155475>

目前您尚未登录，请 [登录](#) 或 [注册](#) 后进行评论

马尔科夫随机场（MRF）与条件随机场（CRF）



jiuerbujie

2015年05月13日 17:53 5232

看《统计学习方法》中的条件随机场一章，觉得应该把马尔科夫随机场和条件随机场一起总结一下。1.马尔科夫随机场1.1无向图假设有一组随机变量 $Y=\{Y_1, Y_2, \dots, Y_n\}$

条件随机场（CRF）



bbbeoy

2018年03月07日 17:36 76

条件随机场应该是机器学习领域比较难的一个算法模型了，难点在于其定义之多（涉及到概率图模型、团等概率）、数学上近似完美（涉及到概率、期望计算，最优化方面的知识），但是其在自然语言处理方面应用效果比较好...

2018，UI设计师如何少走弯路快速提升设计能力！

少走弯路快速提升诀窍：



加入CSDN，享受更精准的内容推荐，与500万程序员共同成长！

登录

注册

条件随机场 (CRF)

a819825294

2016年12月28日 11:15

10097

条件随机场应该是机器学习领域比较难的一个算法模型了，难点在于其定义之多（涉及到概率图模型、团等概率）、数学上近似完美（涉及到概率、期望计算，最优化方面的知识），但是其在自然语言处理方面应用效果比较好，...

机器学习之条件随机场CRF一点理解

FX677588

2016年11月14日 12:51

2397

1. 前言 最近看了一些有关于CRF的论文，基础懂，但是到求解的部分有些疑惑。CRF问题容易构成NP-hard问题，求解过程还需要再学习。下面稍微介绍一些CRF的学习吧！这里前面CRF内容主要参...

条件随机场(CRF) - 3 - 概率计算

xueyingxue001

2016年05月25日 15:50

2226

声明： 1，本篇为个人对《2012.李航统计学习方法.pdf》的学习总结，不得用作商用，欢迎转载，但请注明出处（即：本帖地址）。 2，由于本人在学习初始时有很...

一秒创造无法计算的价值

阿里云

每满2000返200，最高返5000元代金券

CRF-条件随机场

u012771351

2016年11月14日 07:56

436

注：本文中所有公式和思路来自于邹博先生的《机器学习升级版》，我只是为了加深记忆和理解写的本文。这次要分享的是模型是条件随机场模型，听起来就是不是十分的容易，不过也确实是这样的，学习条件随...

条件随机场

qq_20602929

2016年05月22日 23:13

1071

条件随机场(CRF)是给定一组输入随机变量条件下另一组输出随机变量的条件概率分布模型，其特点是假设输出随机变量构成马尔科夫随机场。主要讲述线性链条件随机场，这时，问题变成了由输入序列对输出序列预测...

CRF-条件随机场

haimengao

2014年04月19日 19:58

565

参见：http://blog.csdn.net/chlele0105/article/details/14897761 在组块方面效果还不错

机器学习笔记（八）之条件随机场

yuih344

2018年01月18日 15:55

247

条件随机场（conditional random field，CRF）是给定一组输入随机变量条件下另一组输出随机变量的条件概率分布模型，其特点是假设输出随机变量构成马尔可夫随机场。条件随机场可以用于不...

条件随机场

rosefun96

2017年12月03日 15:48

133

1、引言条件随机场，conditional random field，CRF。 2.

马尔可夫随机场 MRF

pipisorry

2017年10月30日 19:45

2092

马尔可夫网马尔科夫网是使用无向图描述的图模型，是刻画X上联合分布的一种方法，表示一个分解方式,也表示一组条件独立关系。马尔科夫随机场(Markov random field，MRF),也被称为马...

程序员不会英语怎么办？

老司机教你一个数学公式秒懂天下英语

推导条件随机场参数估计的全过程

liujianfei526

2016年02月13日 12:01

656

著作权归作者所有。 商业转载请联系作者获得授权，非商业转载请注明出处。 作者：宁远成梁 链接：http://www.zhihu.com/question/20380549/answer/4506678...

条件随机场(CRF)的一篇不错ppt

jollyjumper

2014年06月22日 13:08

38297

http://wenku.baidu.com/link?url=7LBbXiKPWAPnqYexmBOhz4iCUSny6Ayg3M53Ls0liVKdQlq-9YPNAiW3WKJ5UgihjWKm...

重庆旅游攻略

联系我们

请扫描二维码联系

webmaster@

400-660-011

QQ客服

关于 招聘 广告服务

©1999-2018 CSDN版权所有

京ICP证09002463号

经营性网站备案信息

网络110报警服务

中国互联网举报中心

北京互联网违法和不良信息举报中心