

15

CRF条件随机场的原理、例子、公式推导和应用

 **刘启林**  
国防科学技术大学 软件工程硕士

关注

270 人赞同了该文章

条件随机场（Conditional Random Field，CRF）是自然语言处理的基础模型，广泛应用于中文分词、命名实体识别、词性标注等标注场景。

条件随机场CRF与深度学习结合，产生了BiLSTM-CRF、BiLSTM-CNN-CRF等模型，在中文分词、命名实体识别、词性标注也取得不错的效果。

条件随机场CRF与Attention机制结合，又发展成了Transformer-CRF、BERT-BiLSTM-CRF等模型，使中文分词、命名实体识别、词性标注效果又有显著提高。

本文先引出条件随机场CRF的场、随机场、团、最大团等相关基础概念；接着介绍CRF的原理，重点阐述了线性链条件随机场的原理；然后介绍CRF在中文分词、命名实体识别、词性标注的具体应用；最后对CRF进行总结，指出图模型之间的演化关系，及CRF模型的发展简史。

本文结构如下：

一、CRF基础

 赞同 270

 44 条评论

 分享

 喜欢

 收藏

 申请转载



知乎

首发于
统计学习方法-笔记

4、无向图的因子分解

二、CRF原理

- 1、条件随机场
- 2、线性链条件随机场
- 3、线性链条件随机场公式
- 4、条件随机场例子

三、CRF应用

- 1、中文分词
- 2、命名实体识别
- 3、词性标注

四、CRF总结

- 1、CRF的概括总结
- 2、图模型之间的关系
- 3、CRF的发展简史

直接上PPT。

条件随机场的原理、例子、推导和应用

刘启林 @ 2020.06.11

知乎 @刘启林

CRF条件随机场的原理、例子、公式推导和应用

▲ 赞同 270 ▼ ● 44 条评论 ➦ 分享 ♥ 喜欢 ★ 收藏 📄 申请转载 ...

知乎

首发于
统计学习方法-笔记

目录



一、CRF基础

二、CRF原理

三、CRF应用

四、CRF总结

- 1. 无向图
- 2. 马尔可夫随机场
- 3. 最大团
- 4. 无向图的因子分解

知乎 @刘启林

CRF基础的目录

1、无向图

什么是图？什么是无向图？

1. 无向图

图

> 图的三要素：顶点、边、关联函数

图 G 是指一个有序三元组 $(V(G), E(G), \psi_G)$ ，其中：

- $V(G)$ 是非空的集合， $V(G)$ 中的元素称为图的**顶点**；
- $E(G)$ 是不与 $V(G)$ 相交的集合， $E(G)$ 中的元素称为图的**边**；
- ψ_G 是**关联函数**，它使得每条边对应于 G 的无序顶点对（不必相异）。

无向图

无向图是指边没有方向的图。

> 通常把无向图简化记作 $G=(V, E)$

知乎 @刘启林

什么是图？什么是无向图？

▲ 赞同 270 ▼

● 44 条评论

➤ 分享

♥ 喜欢

★ 收藏

📄 申请转载

...

2. 马尔可夫随机场

场 (field)

➤ 注意：与域(field)的区别

在空间某一区域内，除个别点外，如果对于该区域的每一点 P 都定义了一个确定的量 $f(P)$ ，**该区域**就称为 $f(P)$ 的场。

随机过程

➤ 可以理解为：随时间演变的随机现象

设 T 是一无限实数集，把依赖于**参数** $t \in T$ 的一族（无限多个）随机变量称为随机过程，记为

$$\{X(t), t \in T\}$$

知乎 @刘启林

什么是场？什么是随机过程？

什么是随机场？什么是马尔可夫随机场？

随机场

➤ 从平面(随机过程)到向量空间(随机场)

若 T 是 n 维空间的某个子集，即 t 是一个 n 维向量，此时随机过程又称为随机场。

马尔可夫随机场 (概率无向图模型)

➤ 马尔可夫随机场 = 概率无向图模型

具有马尔可夫性的随机场。

马尔可夫性：

$$P(Y_v | X, Y_w, w \neq v) = P(Y_v | X, Y_w, w \sim v)$$

其中：

$w \sim v$ 表示在图 $G = (V, E)$ 中**与顶点 v 有边连接**的所有顶点 w ；

$w \neq v$ 表示顶点 v 以外的所有顶点；

Y_v 与 Y_w 为顶点 v 与 w 对应的随机变量。

知乎 @刘启林

什么是随机场？什么是马尔可夫随机场？

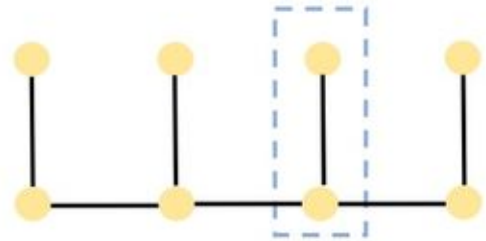
3. 最大团

团 (Clique)

任意两个顶点在图中相邻。

最大团 (Maximal Clique)

若 C 是无向图 G 的一个团，并且不能再加进任何一个 G 的顶点使其成为更大的团。



知乎 @刘启林

什么是团？什么是最大团？

4. 无向图的因子分解

Hammersley-Clifford 定理如下：

4. 无向图的因子分解

概率无向图模型的因子分解

➤ Hammersley-Clifford 定理

给定概率无向图模型，设其无向图为 G ， C 为 G 上的最大团， Y_C 表示 C 对应的随机变量。

那么概率无向图模型的联合概率分布 $P(Y)$ 可写作图中**所有最大团** C 上的函数 $\psi_C(Y_C)$ 的

乘积形似，即

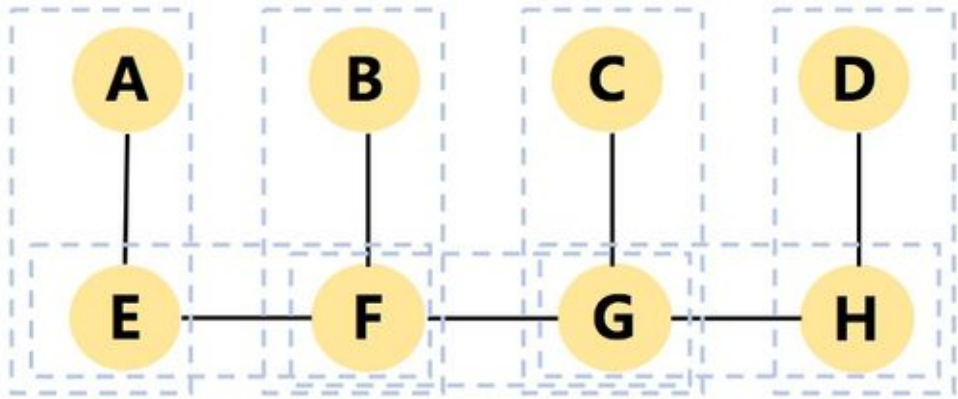
$$P(Y) = \frac{1}{Z} \prod_C \psi_C(Y_C)$$

$$Z = \sum \sum_C \psi_C(Y_C)$$

其中， $\psi_C(Y_C)$ 是 C 上定义的严格正函数，通常取指数函数。

知乎 @刘启林

例1：一个无向图的因子分解例子如下所示。



则由A、E、F、B构成无向图的联合概率分布：

$$P(A, E, F, B) = \frac{1}{Z} e^{f(A,E)} e^{g(E,F)} e^{h(F,B)} = \frac{1}{Z} e^{f(A,E)+g(E,F)+h(F,B)}$$

知乎 @刘启林

无向图的因子分解的例子

二、CRF原理

目录

| | |
|---------|--|
| 一、CRF基础 | 1. 条件随机场 2. 线性链条件随机场 3. 线性链条件随机场公式 4. 条件随机场例子 |
| 二、CRF原理 | |
| 三、CRF应用 | |
| 四、CRF总结 | |

知乎 @刘启林

知乎

首发于
统计学习方法-笔记

2001年, John Lafferty, Andrew McCallum 和 Fernando Pereira, 在论文《Conditional Random fields : Probabilistic Models for Segmenting and Labeling Sequence Data》提出条件随机场。

Conditional Random fields : Probabilistic Models
for Segmenting and Labeling Sequence Data

citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.2...

提出条件随机场。

1. 条件随机场

2001年, John Lafferty, Andrew McCallum 和 Fernando Pereira, 在论文《Conditional Random fields : Probabilistic Models for Segmenting and Labeling Sequence Data》提出条件随机场。

Definition. Let $G = (V, E)$ be a graph such that $\mathbf{Y} = (\mathbf{Y}_v)_{v \in V}$, so that \mathbf{Y} is indexed by the vertices of G . Then (\mathbf{X}, \mathbf{Y}) is a **conditional random field** in case, when conditioned on \mathbf{X} , the random variables \mathbf{Y}_v obey the Markov property with respect to the graph:

$p(\mathbf{Y}_v | \mathbf{X}, \mathbf{Y}_w, w \neq v) = p(\mathbf{Y}_v | \mathbf{X}, \mathbf{Y}_w, w \sim v)$, where $w \sim v$ means that w and v are neighbors in G .

条件随机场定义

If the graph $G = (V, E)$ of \mathbf{Y} is a tree (of which a chain is the simplest example), its cliques are the edges and vertices. Therefore, by the fundamental theorem of random fields (Hammersley & Clifford, 1971), the joint distribution over the label sequence \mathbf{Y} given \mathbf{X} has the form:

$$p_{\theta}(\mathbf{y} | \mathbf{x}) \propto \exp \left(\sum_{e \in E, k} \lambda_k f_k(e, \mathbf{y}|_e, \mathbf{x}) + \sum_{v \in V, k} \mu_k g_k(v, \mathbf{y}|_v, \mathbf{x}) \right) \quad (1)$$

where $\mathbf{y}|_S$ is the set of components of \mathbf{y} associated with the vertices in subgraph S .

Hammersley-Clifford 定理 @刘启林

条件随机场

条件随机场定义如下:

设 X 与 Y 是随机变量， $P(Y|X)$ 是在给定 X 的条件下 Y 的条件概率分布。

若随机变量 Y 构成一个由无向图 $G = (V, E)$ 表示的马尔可夫随机场，即

$$P(Y_v|X, Y_w, w \neq v) = P(Y_v|X, Y_w, w \sim v)$$

对任意顶点 v 成立，则称条件概率分布 $P(Y|X)$ 为条件随机场。

➤ 条件随机场的一般定义

其中：
 $w \sim v$ 表示在图 $G = (V, E)$ 中与顶点 v 有边连接的所有顶点 w ；
 $w \neq v$ 表示顶点 v 以外的所有顶点；
 Y_v 与 Y_w 为顶点 v 与 w 对应的随机变量。

知乎 @刘启林

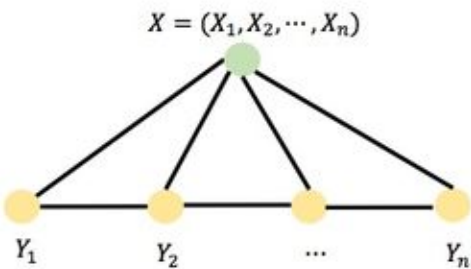
条件随机场定义

2、线性链条件随机场

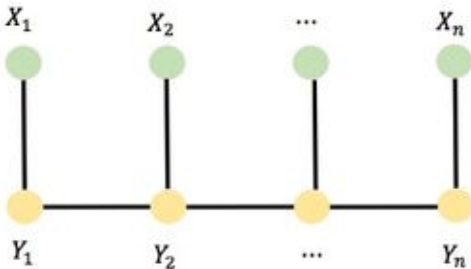
2. 线性链条件随机场

实际应用中，一般设 X 和 Y 有相同的图结构，以线性链条件随机场应用最为广泛。

两种主要线性条件随机场的图结构如下：



最大团是相邻3个顶点的集合



最大团是相邻2个顶点的集合

知乎 @刘启林

线性链条件随机场

线性链条件随机场的定义如下：

若在**给定**随机变量序列 **x** 的条件下，

随机变量序列 **Y** 的条件概率分布 **$P(Y|X)$** 构成条件随机场，即满足马尔可夫性

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

则称 **$P(Y|X)$** 为线性链条件随机场。

其中 **$i = 1, 2, \cdots, n$** ，在 **$i = 1$** 和 **n** 时只考虑单边。

➤ 习惯把“线性链条件随机场”简称为条件随机场(CRF)

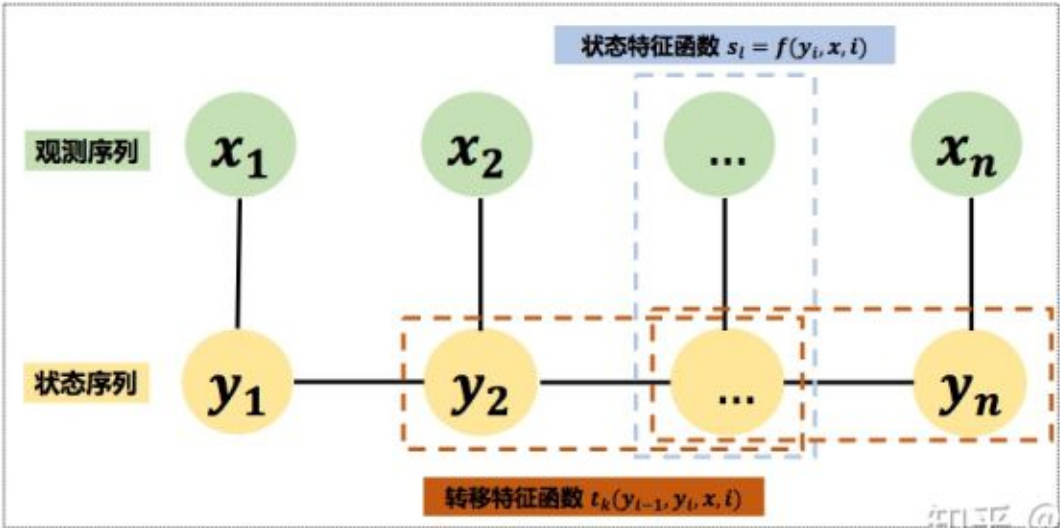
知乎 @刘启林

线性链条件随机场的定义

线性链条件随机场CRF的图结构

线性链条件随机场CRF的图结构

输入：观测序列(x)，输出：状态序列(Y)。



知乎 @刘启林

线性链条件随机场CRF的图结构

3、线性链条件随机场公式

知乎

首发于
统计学习方法-笔记

设 $P(Y|X)$ 为线性链条件随机场，则在随机变量 X 取值为 x 的条件下，随机变量 Y 取值为 y 的条件概率具有如下形式：

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

➤ 表示给定输入序列 x ，对输出序列 y 预测的条件概率

其中：

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

t_k, s_l ：特征函数

λ_k, u_l ：对应权值

$Z(x)$ ：归一化因子

知乎 @刘启林

线性链条件随机场公式

特征函数定义如下：

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

t_k, s_l 特征函数是二值函数，函数值为0或者1——满足特征条件时取值为1，否则为0；

t_k 是定义在**边**上的特征函数，称为转移特征，依赖于**当前和前一个位置**；

$$t_k(y_{i-1}, y_i, x, i) = \begin{cases} 1, & \text{条件} \\ 0, & \text{其他} \end{cases}$$

s_l 是定义在**节点**上的特征函数，称为状态特征，依赖于**当前位置**；

$$s_l(y_i, x, i) = \begin{cases} 1, & \text{条件} \\ 0, & \text{其他} \end{cases}$$

知乎 @刘启林

特征函数

为了简单起见，将转移特征和状态特征及其权值用统一符号表示。条件随机场简化公式如下：

对公式 (1) 进行简化，得：

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

(2)

其中：

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

知乎 @刘启林

条件随机场简化公式

4、条件随机场例子

例2：已知中文文本"刘启林"，采用IOB标注方法，求实体标注序列为{ B, I, I }的概率。

4. 条件随机场例子

例2：已知中文文本"刘启林"，采用IOB标注方法，求实体标注序列为{ B, I, I }的概率。

分析：

输入观测序列 $X = (X_1, X_2, X_3)$ ，即 $x_1 = \text{刘}, x_2 = \text{启}, x_3 = \text{林}$ 。

输出标注序列 $Y = (Y_1, Y_2, Y_3)$ ，则 y_1, y_2, y_3 取值{I, O, B}。

解：

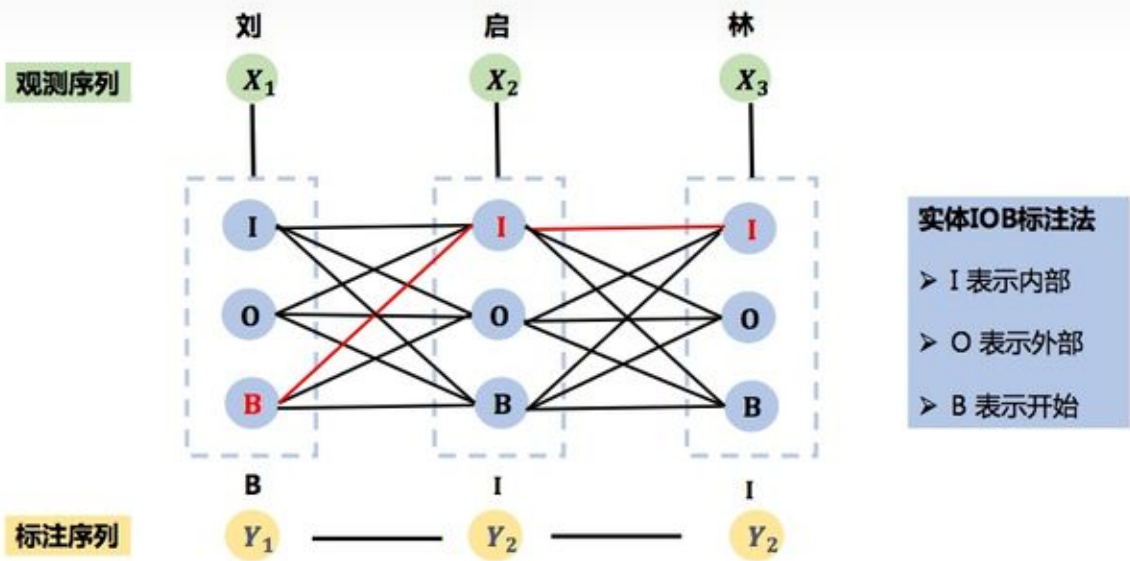
假设其它特征函数为0，则：

$$P(y_1, y_2, y_3|x) = \frac{1}{Z(x)} (1 + 0.5 + 1 + 0.5 + 0.5)$$

| 特征函数 | 函数条件 | 函数值 | 权重 |
|-------|--------------------|-----|-----|
| t_1 | $y_1 = B, y_2 = I$ | 1 | 1 |
| | 其他 | 0 | |
| t_2 | $y_2 = I, y_3 = I$ | 1 | 0.5 |
| | 其他 | 0 | |
| s_1 | $y_1 = B$ | 1 | 1 |
| | 其他 | 0 | |
| s_2 | $y_2 = I$ | 1 | 0.5 |
| | 其他 | 0 | |
| s_3 | $y_3 = I$ | 1 | 0.5 |
| | 其他 | 0 | |

知乎 @刘启林

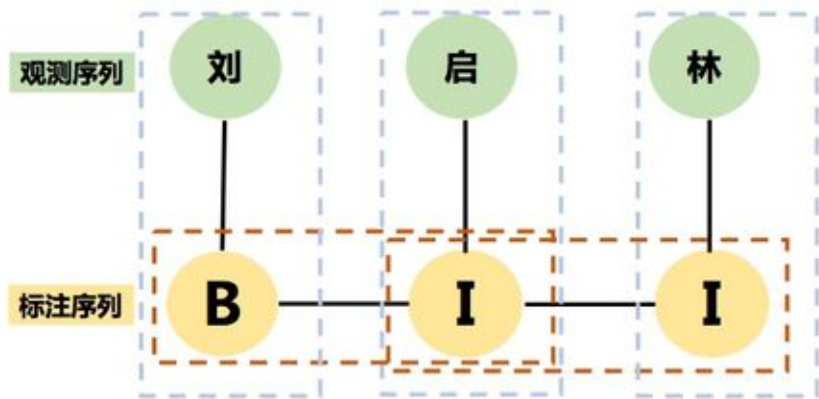
求实体标注序列为{ B, I, I }的概率



知乎 @刘启林

例1的几何描述

例1的标注序列为{B、I、I}的联合概率分布如下：



则“刘启林”的标注序列为{B、I、I}的联合概率分布：

$$\begin{aligned} P(B, I, I) &= \frac{1}{Z} e^{f(\text{刘}, B)} e^{g(B, I)} e^{f(\text{启}, I)} e^{g(I, I)} e^{f(\text{林}, I)} \\ &= \frac{1}{Z} e^{f(\text{刘}, B) + g(B, I) + f(\text{启}, I) + g(I, I) + f(\text{林}, I)} \end{aligned}$$

知乎 @刘启林

例1的标注序列为{B、I、I}的联合概率分布

目 录

一、CRF基础

二、CRF原理

三、CRF应用

四、CRF总结

1. 中文分词

2. 命名实体识别

3. 词性标注

知乎 @刘启林

CRF应用的目录

1、中文分词

基于CRF由字构词方法的基本思想，基本原理如下：

基于CRF中文分词的基本思想，基本原理

CRF中文分词的图结构如下：

CRF中文分词的图结构

2、命名实体识别

基于CRF的命名实体识别过程如下：

基于CRF的命名实体识别

CRF命名实体识别的图结构如下：

一、CRF概述

基于CRF词性标注方法的基本思想，基本原理如下：

基于CRF词性标注的基本思想，基本原理

CRF中文词性标注的图结构如下：

四、CRF总结

CRF总结的目录

1、CRF的概括总结

CRF的概括总结

2、图模型之间的关系

朴素贝叶斯、HMM、逻辑回归、CRF等图模型关系如下：

朴素贝叶斯、HMM、逻辑回归、CRF等图模型关系

朴素贝叶斯、HMM、逻辑回归、CRF对比如下表所示：

朴素贝叶斯、HMM、逻辑回归、CRF对比

更多HMM可参考：

更多LR逻辑回归模型可参考：

3、CRF的发展简史

CRF的发展简史

机器学习阶段：CRF

深度学习阶段：BiLSTM-CRF、BiLSTM-CNN-CRF

Attention阶段：Transformer-CRF、BERT-BiLSTM-CRF

条件随机场CRF一直是标注问题的基础模型。

由于能力和水平的限制，我的可能是错的。

参考文献：

- 1、王元等, 数学大词典(第二版), 科学出版社[M], 2017.09
- 2、John Lafferty, Andrew McCallum, and Fernando Pereira, "Conditional Random Fields:

4、宗成庆, 统计自然语言处理(第2版), 清华大学出版社[M], 2013.08

5、Charles Sutton and Andrew McCallum, An Introduction to Conditional Random Fields, 2011

编辑于 2020-11-17 01:11




「稀罕作者」

赞赏

还没有人赞赏，快来当第一个赞赏的人吧！

条件随机场 CRFs 自然语言处理

文章被以下专栏收录

- 统计学习方法-笔记
- python遇见NLP
主要关注机器学习、自然语言处理 分享，让知识共享！
- 数据科学之路
数据驱动的都是数据科学

推荐阅读

全方位理解条件随机场
(CRF) · 原理 应用案例

一文理解条件随机场CRF

一前

知乎

首发于

统计学习方法-笔记

<https://www.guandun.com/p/3373>

<https://www.cnblogs.com/pangxi>

闫冬

发表于机器学习

序列标注 (sequence labeling)

智勇香肠番茄怪

黄

44 条评论

切换为时间排序

评论区功能升级中



精选评论 (1)

刘启林 (作者) 回复 phoenix bai

2020-11-18

语言天然具有先后顺序，成线性特性。比如：一句话“我喜欢北京”，X输入就是“我喜欢北京”。输出Y：“我 喜欢 北京”（分词）

3 查看回复

评论 (44)

- 君夕夕

06-05

你好，那个中文分词的原理可以讲解下吗？

1
- 刘启林 (作者) 回复 君夕夕

06-05

中文分词的原理见另一篇文章：中文分词的原理、方法与工具

赞
- 再见孙悟空

03-16

感谢，有图有例子非常好理解

1
- 刘启林 (作者) 回复 再见孙悟空

03-16

谢谢你的支持与肯定，一起交流学习。

赞
- 如果

11-27

太强了吧

知乎

首发于
统计学习方法-笔记

carney

11-18

您好，我觉得你这个讲的特别好，非常感谢您！请问能麻烦您分享我一份PPT么？邮箱：1478915252@qq.com。非常感谢非常感谢

赞



王宇飞

10-11

这是我看到最清晰的解读了，非常感谢您！请问能麻烦您分享我一份PPT么？邮箱：654775341@qq.com。非常感谢非常感谢

赞



zhibo

05-19

感谢，解释的很清楚

赞



王攀

05-17

刘老师您好，请问能否分享一下ppt?sui6662012@163.com

赞



德乙球队沙尔克

05-05

谢谢你，终于搞懂了 这个势函数是什么寄吧

赞



呗织哩

04-27

做得太好了，可以分享一下ppt吗？

邮箱是 chenyuqian111@163.com 谢谢~

赞



Ben

04-25

MRF和CRF的区别是什么呢?看PPT中两者的数学公式都是一样的

赞



风口吹风 回复 Ben

10-12

MRF里面应该还是没有X的，作者应该是写错了

赞



嗯嗯

04-01

我也想要一份PPT, 1934234962@qq.com, 谢谢!

👍 赞



lunature

02-19

关于随机过程的定义: 随机变量是随机过程?变量是过程?总感觉有点不太对

👍 赞



刘启林 (作者) 回复 lunature

02-19

是一簇随机变量。

标准定义可以参阅王元院士的《数学大词典(第二版)》。

👍 赞



西伯利亚寒流

01-15

能否分享一下PPT, jacob081.com。谢谢!

👍 赞



是念

2020-12-30

我也想要一份PPT, 谢谢, 476226078@qq.com

👍 赞



CheonHye

2020-12-03

例2: 已知中文文本"刘启林", 采用IOB标注方法, 求实体标注序列为{ B, I, I }的概率。在求解的时候并没有用到"刘启林"呀? 那是否是说任意三个字最后的结果都是一样的?

👍 赞



刘启林 (作者) 回复 CheonHye

2020-12-05

我的举例(例2)其实是文本的序列组合的一种, 只不过它是以字组成序列的, 但是也可以以词组成序列的, 不同场景不同算法不同选择而已。例2的"刘启林"按IOB方法组合(以字为单位), 理论上有27种组合方法, 但例子是求BII这个组合序列的概率。初始化概率可以设置一样, 但是训练后概率肯定不一样。

举一个实际应用的例子, 比如命名实体识别方法中, 第一步可以切字, 这时就没有用到"刘启林", 但是也可以切词, 这时就会用到的"刘启林"。

知乎

首发于
统计学习方法-笔记

CheonHye 回复 刘启林 (作者)

2020-12-07

非常感谢您如此详尽的回复!!!

👍 赞



五车

2020-12-02

答主没有说清楚crf如何训练。对比代码中crf训练，差距很大的。这有点误导读者。不过资料引用挺全面。给个赞。

👍 赞



王雅宁

2020-11-22

能否贴一个ppt 的在线下载链接呢，或者发我邮箱一份，非常感谢 shiter@live.cn

👍 赞



phoenix bai

2020-11-18

请教: 线性链条件随机场的定义中, 怎么定义X序列和Y序列是线性表示的随机变量序列? 这里指的线性特性是什么?

👍 赞



刘启林 (作者) 回复 phoenix bai

2020-11-18

语言天然具有先后顺序，成线性特性。比如：一句话“我喜欢北京”，X输入就是“我喜欢北京”。输出Y：“我 喜欢 北京”（分词）

👍 3 查看回复



phoenix bai 回复 phoenix bai

2020-11-18

get, 谢谢

👍 赞

1 2 下一页