

## klzhang的专栏

没有方向，我便不会行走。。 Read More , Write More , Think More。。。

[目录视图](#)[摘要视图](#)[RSS 订阅](#)

### 个人资料



zkl99999

[关注](#)[发私信](#)

访问： 613463次

积分： 7237

等级： **BLOG > 6**

排名： 第3235名

原创： 29篇

转载： 372篇

译文： 1篇

评论： 49条

异步赠书：9月重磅新书升级，本本经典

程序员9月书讯

每周荐书：ES6、虚拟现实、物联网（评论送书）

## 交叉熵 机器学习模型loss指标

2016-04-02 12:56

4977人阅读

评论(0)

收藏

分类：

机器学习 (73)

目录(?)

[+]

## 华夏35度

Data Mining <http://www.cnblogs.com/zhangchaoyang/article/details/50483212>

utm\_source=tuicool

菜鸟从零开始学习Deep learning (81) <http://blog.csdn.net/yihaizhiyan/article/details/50483212>

### 信息论

信息论 (Information Theory) 是概率论与数理统计的一个分枝。用于信息处理、信息熵、通信系统、数据传输、率失真理论、密码学、信噪比、数据压缩和相关课题。

## 基本概念

先说明一点：在信息论里面对数log默认都是指以2为底数。

关闭

文章搜索

文章分类

- 机器学习 (74)
- Python (23)
- 统计学习方法 (40)
- 三维重建 (50)
- PCL (34)
- OpenCV (55)
- Linux/Unix (26)
- OpenCL (3)
- face (4)
- 编程所预问题 (4)
- Unity3D (10)
- 数据结构与算法 (31)
- C++/C (75)
- 生活感悟 (28)
- window (22)
- openGL (21)
- Hadoop (7)
- Kinect (16)
- Matlab (1)
- Deep Learning (6)
- ROS (8)
- CUDA (5)
- CG (1)

自信息量



联合自信息量



条件自信息量



信息熵



条件熵



联合熵



根据链式规则，有



可以得出



信息增益Information Gain

系统原先的熵是 $H(X)$ ，在条件 $Y$ 已知的情况下系统的熵（条件熵）为 $H(X|Y)$ ，信息增益就是这两个熵的差值。



熵表示系统的不确定度，所以信息增益越大表示条件 $Y$ 对于确定系统的贡献越大。

信息增益在特征选择中的应用

由（7）式可以直接推出词条 $w$ 的信息增益，（7）式中的 $X$ 代表类别的集合， $Y$ 代表 $w$ 存在和不存在两种情况



关闭

互联网 (37)

O2O (3)

推荐系统 (2)

HTML5 (1)

JS/Java (3)

AR/VR (13)

NLP (3)

Docker (1)

TensorFlow (6)

面试 (9)

软件架构设计 (1)

QT5 (2)

文章存档

2017年09月 (1)

2017年03月 (2)

2017年02月 (1)

2016年12月 (1)

2016年09月 (1)

展开

阅读排行

caffe 中的损失函数分析以及lo... (12046)

张正友平面标定方法 超详细 (10846)

神经网络理论 超全超详细 通熟.. (9425)

人体行为识别特征点提取 综述... (9202)

点云三维重建有关原理 (8621)



$p(c_i)$ 是第*i*类文档出现的概率； $p(w)$ 是在整个训练集中包含*w*的文档占全部文档的比例； $p(c_i|w)$ 表示出现*w*的文档集合中属于类别*i*的文档所占的比例；表示没有出现*w*的文档集合中属于类别*i*的文档所占的比例。

信息增益在决策树中的应用

outlook	temperature	humidity	windy	play
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
overcast	hot	high	FALSE	yes
rainy	mild	high	FALSE	yes
rainy	cool	normal	FALSE	yes
rainy	cool	normal	TRUE	no
overcast	cool	normal	TRUE	yes
sunny	mild	high	FALSE	no
sunny	cool	normal	FALSE	yes
rainy	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no

（7）式中的X表示打球和不打球两种情况。

只看最后一列我们得到打球的概率是9/14，不打球的概率是5/14。因此在没有任何先验信息的情况下，系统的熵（不确定性）为



--	--	--	--	--

关闭

- 从卡尔曼滤波到粒子滤波 很详... (8222)
- 28款GitHub最流行的开源机器... (6735)
- LSD-SLAM深度学习 ( 1 ) -基... (6331)
- 用hector mapping构建地图 (6064)
- 斯坦福CS231N深度学习与计算.. (5337)

评论排行

- Kinect深度图与摄像头RGB的... (10)
- Structure from Motion Using Op... (6)
- 从卡尔曼滤波到粒子滤波 很详... (3)
- pcl c++ 文件后缀名批处理修改... (3)
- 视觉SLAM漫谈123456 (2)
- PCL 室内三维重建 (2)
- 张正友平面标定方法 超详细 (2)
- Bouguet极线校正的方法 (2)
- 概念：malloc/free与new/delete... (1)
- Visual Studio 2010建模学习(十... (1)

推荐文章

- \* CSDN新版博客feed流内测用户征集令
- \* Android检查更新下载安装
- \* 动手打造史上最简单的 Recycleview 侧滑菜单
- \* TCP网络通讯如何解决分包粘包问题
- \* SDCC 2017之大数据技术实战线上峰会
- \* 快速集成一个视频直播功能

最新评论

outlook			temperature			humidity			windy			play	
	yes	no		yes	no		yes	no		yes	no	yes	no
sunny	2	3	hot	2	2	high	3	4	FALSE	6	2	9	5
overcast	4	0	mild	4	2	normal	6	1	TRUR	3	3		
rainy	3	2	cool	3	1								

如果选outlook作为决策树的根节点，（ 7 ）式中的Y为集合{sunny、overcast、rainy}，此时的条件熵为



即选择outlook作为决策树的根节点时，信息增益为0.94-0.693=0.247。

同样方法计算当选择temperature、humidity、windy作为根节点时系统的信息增益，选择IG值最大的作为最优特征根节点。

关闭

互信息Mutual Informantion

y<sub>j</sub>对x<sub>i</sub>的互信息定义为后验概率与先验概率比值的对数。



互信息越大，表明y<sub>j</sub>对于确定x<sub>i</sub>的取值的贡献度越大。

系统的平均互信息



可见平均互信息就是信息增益！

一种基于OpenCV的三维重建实现方案

sg2017918 : 没多大用

张正友平面标定方法 超详细

feiyang6064 : 张正友相机标定对于已知的图像没有标定物怎么处理呀

蒙特卡洛树搜索 MCTS

snailYWW : 你好，我想问下，蒙特卡洛树搜索 里面有包含 minmax么？

Kinect深度图与摄像头RGB的标定与配准

zhurui1230 : 你好，刚开始接触Kinect，我想问一下，不使用第三方摄像头，使用Kinect的rgb和深度图还需要...

Structure from Motion Using OpenCV

sinat\_36486976 : 请问博主，完整的章节在哪里可以看到

用Tensorflow基于Deep Q Learning DQN 玩...

sinat\_32456781 : 你好，楼主，我把代码运行了一下，为啥每次都会碰到水管

Structure from Motion Using OpenCV

znlangel : @u013472489:您好，我也刚好看到书上的SFM项目了，但是有些配置问题想问下您是怎么解决的，...

Structure from Motion Using OpenCV

znlangel : @sinat\_32274457:您好，我也刚好看到书上的SFM项目了，但是有些配置问题想问下您是怎么...

Structure from Motion Using OpenCV

znlangel : @sinat\_32274457:您好，请问，你的这个重构的项目可以运行成功吗，因为书有直接配套的源码...

Structure from Motion Using OpenCV

sinat\_32274457 : 楼主，想问一下，我采用的global 的方法，即求得的R与t都是全局坐标系下，分别代表相机的方向与位...

## 互信息在特征选择中的应用

词条 $w$ 与类别 $c_i$ 的互信息为



$p(w)$ 表示出现 $w$ 的文档点总文档数目的比例， $p(w|c_i)$ 表示在类别 $c_i$ 中出现 $w$ 的文档点总文档数目的比例。

对整个系统来说，词条 $w$ 的互信息为



最后选互信息最大的前 $K$ 个词条作为特征项。

## 交叉熵Cross Entropy

交叉熵是一种万能的Monte-Carlo技术，常用于稀有事件的仿真建模、多峰函数的最优化问题。交叉熵经典的旅行商问题、背包问题、最短路问题、最大割问题等。这里给一个文章链接：[A Tutorial on the Cross Entropy Method](#)

交叉熵算法的推导过程中又牵扯出来一个问题：如何求一个数学期望？常用的方法有这么几种：

- 概率方法，比如Crude Monte-Carlo
- 测度变换法change of measure
- 偏微分方程的变量代换法
- Green函数法
- Fourier变换法

关闭

在实际中变量 $X$ 服从的概率分布 $h$ 往往是不知道的，我们会用 $g$ 来近似地代替 $h$ ----**这本质上是一种函数估计**。有一种度量 $g$ 和 $h$ 相近程度的方法叫 Kullback-Leibler距离，又叫交叉熵：



通常选取 $g$ 和 $h$ 具有相同的概率分布类型（比如已知 $h$ 是指数分布，那么就选 $g$ 也是指数分布）----**参数估计**，只是pdf参数不一样（实际上 $h$ 中的参数根本就是未知的）。

## 基于期望交叉熵的特征项选择



$p(c_i|w)$ 表示在出现词条 $w$ 时文档属于类别 $c_i$ 的概率。

交叉熵反应了文本类别的概率分布与在出现了某个词条的情况下文本类别的概率分布之间的距离。词条的交叉熵越大，对文本类别分布影响也就越大。所以选CE最大的K个词条作为最终的特征项。

本文来自:博客园 ( 华夏35度 ) <http://www.cnblogs.com/zhangchaoyang> 作者:Orisun

Suppose you have just three training items with the following computed outputs and target outputs:

computed				target		
-----				-----		
0.1	0.3	0.6		0	0	1
0.2	0.6	0.2		0	1	0
0.3	0.4	0.3		1	0	0

The mean (average) CE error for the three items is the sum of the CE errors divided by the number of items. A fancy way to express CE error with a function is shown in **Figure 2**.

$$-\sum_{i=0}^n \ln(o_i) * t_i$$

Figure 2. Cross Entropy Error Function.

关闭

In words this means, "Add up the product of the log to the base e of each computed output times its corresponding target output, and then take the negative of that sum." So for the three items above, the CE of the first item is  $-(\ln(0.1)*0 + \ln(0.3)*0 + \ln(0.6)*1) = -(0 + 0 - 0.51) = 0.51$ . The CE of the second item is  $-(\ln(0.2)*0 + \ln(0.6)*1 + \ln(0.2)*0) = -(0 - 0.51 + 0) = 0.51$ . The CE of the third item is  $-(\ln(0.3)*1 + \ln(0.4)*0 + \ln(0.3)*0) = -(-1.2 + 0 + 0) = 1.20$ . So the mean cross entropy error for the three-item data set is  $(0.51 + 0.51 + 1.20) / 3 = 0.74$ .

From:

<http://visualstudiomagazine.com/articles/2014/04/01/neural-network-cross-entropy-error.aspx>

顶 踩  
2 0

- [上一篇](#) [神经网络理论 超全超详细 通俗易懂](#)
- [下一篇](#) [Ubuntu 14.04下OpenCV 3.0 安装 测试](#)

#### 相关文章推荐

- [tf.nn.softmax\\_cross\\_entropy\\_with\\_logits\(\)笔记及交叉...](#)
- [决策树系列之三 GBDT \( MART \) 迭代决策树](#)
- [Presto的服务治理与架构在京东的实践与应用--王哲...](#)
- [Retrofit 从入门封装到源码解析](#)
- [交叉熵代价函数 \( 作用及公式推导 \)](#)
- [机器学习基础 \( 五十八 \) —— 香农熵、相对熵](#)
- [深入掌握Kubernetes应用实践--王渊命](#)
- [自然语言处理](#)
- [交叉熵代价函数](#)
- [神经网络与机器学习 \( 原书第3版 \)](#)
- [Python基础知识汇总](#)
- [机器学习笔记 \( 三 \) 朴素贝叶斯分类](#)
- [caffe Sigmoid cross entropy loss 交叉熵损失函数](#)
- [初学者如何查阅自然语言处理 \( NLP \) 领域学术资料](#)
- [Android核心技术详解](#)
- [一种图像去噪算法的实现](#)

[关闭](#)

#### 查看评论

暂无评论


您还没有登录,请[\[登录\]](#)或[\[注册\]](#)

\* 以上用户言论只代表其个人观点，不代表CSDN网站的观点或立场

---

[公司简介](#) | [招贤纳士](#) | [广告服务](#) | [联系方式](#) | [版权声明](#) | [法律顾问](#) | [问题报告](#) | [合作伙伴](#) | [论坛反馈](#)

---

[网站客服](#)   [杂志客服](#)   [微博客服](#)   [webmaster@csdn.net](mailto:webmaster@csdn.net)   400-660-0108 | 北京创新乐知信息技术有限公司 版权所有 | 江苏知之为计算机有限公司 | 江苏乐知信息技术有限公司  
京 ICP 证 09002463 号 | Copyright © 1999-2017, CSDN.NET, All Rights Reserved 

关闭