

电子发烧友论坛 > 小组 >> 技术 > 人工智能 / 非常通俗的朴素贝叶斯算法（Naive Bayes） ...

验] 非常通俗的朴素贝叶斯算法（Naive Bayes）

zpwsmile 2018-10-8 10:14:31 8160 贝叶斯分类器 朴素贝叶斯 贝叶斯公式 分享

三

版块导航

简化多通道频率和相位同步  
加速开发过程 - 简化无线系统  
2020年1月7日 10:00AM  
免费报名

朴素贝叶斯算法（Naive Bayes）

阅读目录

- 一、病人分类的例子
- 二、朴素贝叶斯分类器的公式
- 三、账号分类的例子
- 四、性别分类的例子

生活中很多场合需要用到分类，比如新闻分类、病人分类等等。  
本文介绍朴素贝叶斯分类器（Naive Bayes classifier），它是一种简单有效的常用分类算法。

1 病人分类的例子让我从一个例子开始讲起，你会看到贝叶斯分类器很好懂，一点都不难。

某个医院早上收了六个门诊病人，如下表截图。

症状	职业	疾病
打喷嚏	护士	感冒
打喷嚏	农夫	过敏
头痛	建筑工人	脑震荡
头痛	建筑工人	感冒
打喷嚏	教师	感冒
头痛	教师	脑震荡

现在又来了第七个病人，是一个打喷嚏的建筑工人。请问他患上感冒的概率有多大？

根据贝叶斯定理：

$$P(A|B) = P(B|A) P(A) / P(B)$$

可得：

$$\begin{aligned} &P(\text{感冒}|\text{打喷嚏}\times\text{建筑工人}) \\ &= P(\text{打喷嚏}\times\text{建筑工人}|\text{感冒}) \times P(\text{感冒}) \\ &\quad / P(\text{打喷嚏}\times\text{建筑工人}) \end{aligned}$$

假定"打喷嚏"和"建筑工人"这两个特征是独立的，因此，上面的等式就变成了：



人工智能

145个成员聚集在这个小组

加入小组

最新话题

- 【瑞芯微RK1808计算棒试用体验】2、RK1808被动模式下mobilenet\_v1模型测试
- 【瑞芯微RK1808计算棒试用体验】如何获取识别后的信息
- 【瑞芯微RK1808计算棒试用体验】TB-RK1808S0试用报告-----来自底层驱动软件工程师
- 【瑞芯微RK1808计算棒试用体验】--计算棒的使用效果（1）环境搭建及设置
- 【瑞芯微RK1808计算棒试用体验】（1）-----开发环境搭建和mobilenet\_v1示例体验

热门话题

- 初学AI人工智能需要哪些技术？这几本书为你解答  
1952浏览 18评论
- Python人工智能学习工具包+入门与实践资料集锦  
3148浏览 7评论
- AI人工智能：零基础入门机器人开发教程  
2804浏览 4评论
- 【Toybrick RK3399Pro AI开发板试用体验】Toybrick RK3399Pro安装ROS系统  
870浏览 4评论
- 人工智能语音芯片行业的发展趋势如何？  
639浏览 4评论

创建小组步骤

站长推荐 4/6 关闭

【开发板活动第1期】电子发烧友“开发板交换站”活动上线，一起来换积分！换板！换礼...

本次活动为各位热忠开发板研究设计的发烧友量身打造，用你的闲置开发板来兑换成积分并可置换你想要的开发板或礼品！

查看 >

[创建属于自己的小组](#)

三  
版  
块  
导  
航

$$\begin{aligned}
 &P(\text{感冒}|\text{打喷嚏}\times\text{建筑工人}) \\
 &= P(\text{打喷嚏}|\text{感冒}) \times P(\text{建筑工人}|\text{感冒}) \times P(\text{感冒}) \\
 &\quad / P(\text{打喷嚏}) \times P(\text{建筑工人})
 \end{aligned}$$

这是可以计算的：

$$\begin{aligned}
 &P(\text{感冒}|\text{打喷嚏}\times\text{建筑工人}) \\
 &= 0.66 \times 0.33 \times 0.5 / 0.5 \times 0.33 \\
 &= 0.66
 \end{aligned}$$

因此，这个打喷嚏的建筑工人，有66%的概率是得了感冒。同理，可以计算这个病人患上过敏或脑震荡的概率。

比较这几个概率，就可以知道他最可能得什么病。

这就是贝叶斯分类器的基本方法：在统计资料的基础上，依据某些特征，计算各个类别的概率，从而实现分类。

## 2 朴素贝叶斯分类器的公式

假设某个体有n项特征（Feature），分别为F1、F2、...、Fn。现有m个类别（Category），分别为C1、C2、...、Cm。

贝叶斯分类器就是计算出概率最大的那个分类，也就是求下面这个算式的最大值：

$$\begin{aligned}
 &P(C|F_1F_2...F_n) \\
 &= P(F_1F_2...F_n|C)P(C) / P(F_1F_2...F_n)
 \end{aligned}$$

由于  $P(F_1F_2...F_n)$  对于所有的类别都是相同的，可以省略，问题就变成了求的最大值：

$$P(F_1F_2...F_n|C)P(C)$$

朴素贝叶斯分类器则是更进一步，假设所有特征都彼此独立，因此：

$$\begin{aligned}
 &P(F_1F_2...F_n|C)P(C) \\
 &= P(F_1|C)P(F_2|C) \dots P(F_n|C)P(C)
 \end{aligned}$$

上式等号右边的每一项，都可以从统计资料中得到，由此就可以计算出每个类别对应的概率，从而找出最大概率的那个类。

虽然“所有特征彼此独立”这个假设，在现实中不太可能成立，但是它可以大大简化计算，而且有研究表明对分类结果的准确性影响不大。

下面再通过两个例子，来看如何使用朴素贝叶斯分类器。

站长推荐 4/6

关闭

【开发板活动第1期】电子发烧友“开发板交换站”活动上线，一起来换积分！换板！换礼...

本次活动为各位热衷开发板研究设计的发烧友量身打造，用你的闲置开发板来兑换成积分并可置换你想要的开发板或礼品！

[查看 »](#)

## 3 账号分类的例子

本例摘自张洋的《算法杂货铺——分类算法之朴素贝叶斯分类》。

根据某[社区网站](#)的抽样统计，该站10000个账号中有89%为真实账号（设为C0），11%为虚假账号（设为C1）。

$$C0 = 0.89$$

$$C1 = 0.11$$

接下来，就要用统计资料判断一个账号的真实性。假定某一个账号有以下三个特征：

F1: 日志数量/注册天数

F2: 好友数量/注册天数

F3: 是否使用真实头像（真实头像为1，非真实头像为0）

$$F1 = 0.1$$

$$F2 = 0.2$$

$$F3 = 0$$

请问该账号是真实账号还是虚假账号？

方法是使用朴素贝叶斯分类器，计算下面这个计算式的值。

$$P(F1|C)P(F2|C)P(F3|C)P(C)$$

虽然上面这些值可以从统计资料得到，但是这里有一个问题：F1和F2是连续变量，不适宜按照某个特定值计算概率。

一个技巧是将连续值变为离散值，计算区间的概率。

比如将F1分解成[0, 0.05]、(0.05, 0.2)、[0.2, +∞]三个区间，然后计算每个区间的概率。在我们这个例子中，F1等于0.1，落在第二个区间，所以计算的时候，就使用第二个区间的发生概率。

根据统计资料，可得：

$$P(F1|C0) = 0.5, P(F1|C1) = 0.1$$

$$P(F2|C0) = 0.7, P(F2|C1) = 0.2$$

$$P(F3|C0) = 0.2, P(F3|C1) = 0.9$$

因此：

$$\begin{aligned} &P(F1|C0) P(F2|C0) P(F3|C0) P(C0) \\ &= 0.5 \times 0.7 \times 0.2 \times 0.89 \\ &= 0.0623 \end{aligned}$$

$$\begin{aligned} &P(F1|C1) P(F2|C1) P(F3|C1) P(C1) \\ &= 0.1 \times 0.2 \times 0.9 \times 0.11 \\ &= 0.00198 \end{aligned}$$

可以看到，虽然这个用户没有使用真实头像，但是他是真实账号的概率，比虚假账号高出30多倍，因此判断这个账号为真。

站长推荐 4/6

关闭

【开发板活动第1期】电子发烧友“开发板交换站”活动上线，一起来换积分！换板！换礼...

本次活动为各位热衷开发板研究设计的发烧友量身打造，用你的闲置开发板来兑换成积分并可置换你想要的开发板或礼品！

[查看 »](#)

4 性别分类的例子

本例摘自维基百科，关于处理连续变量的另一种方法。下面是一组人类身体特征的统计资料。

性别	身高（英尺）	体重（磅）	脚掌（英寸）
男	6	180	12
男	5.92	190	11
男	5.58	170	12
男	5.92	165	10
女	5	100	6
女	5.5	150	8
女	5.42	130	7
女	5.75	150	9

已知某人身高6英尺、体重130磅，脚掌8英寸，请问该人是男是女？根据朴素贝叶斯分类器，计算下面这个式子的值。

$$P(\text{身高}|\text{性别}) \times P(\text{体重}|\text{性别}) \times P(\text{脚掌}|\text{性别}) \times P(\text{性别})$$

这里的困难在于，由于身高、体重、脚掌都是连续变量，不能采用离散变量的方法计算概率。而且由于样本太少，所以也无法分成区间计算。怎么办？

这时，可以假设男性和女性的身高、体重、脚掌都是正态分布，通过样本计算出均值和方差，也就是得到正态分布的密度函数。

有了密度函数，就可以把值代入，算出某一点的密度函数的值。

比如，男性的身高是均值5.855、方差0.035的正态分布。

所以，男性的身高为6英尺的概率的相对值等于1.5789（大于1并没有关系，因为这里是密度函数的值，只用来反映各个值的相对可能性）。

$$p(\text{height}|\text{male}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(6-\mu)^2}{2\sigma^2}\right) \approx 1.5789$$

有了这些数据以后，就可以计算性别的分类了。

$$\begin{aligned} &P(\text{身高}=6|\text{男}) \times P(\text{体重}=130|\text{男}) \times P(\text{脚掌}=8|\text{男}) \times P(\text{男}) \\ &= 6.1984 \times e^{-9} \\ &P(\text{身高}=6|\text{女}) \times P(\text{体重}=130|\text{女}) \times P(\text{脚掌}=8|\text{女}) \times P(\text{女}) \\ &= 5.3778 \times e^{-4} \end{aligned}$$

可以看到，女性的概率比男性要高出将近10000倍，所以判断该人为女性。



只看该作者 淘帖 举报

经验

全部 贝叶斯分类器 朴素贝叶斯 贝叶斯公式

用的分类方法：朴素贝叶斯法 47

站长推荐 4/6 关闭

【开发板活动第1期】电子发烧友“开发板交换站”活动上线，一起来换积分！换板！换礼...

本次活动为各位热衷开发板研究设计的发烧友量身打造，用你的闲置开发板来兑换成积分并可置换你想要的开发板或礼品！



查看 »

电子工程师的“翰林院”，你加入了么？  
即刻加入 还有机会 **赢200元手机话费** [立即加入](#)

只有小组成员才能发言，[加入小组>>](#)

华秋（原“华强聚丰”）： 电子发烧友 华秋开发 华秋电路（原“华强PCB”） 华秋商城（原“华强芯城”） 华秋智造

My ElecFans APP

产业	技术	特色栏目	社群	供应链服务	华秋	联系我们	
物联网	可编程逻辑	专栏	问答	项目外包	关于我们	大中华区广告合作	供应链服
机器人	电源/新能源	电子说	论坛	BOM配单	投资关系	Ben Tseng	江良华
人工智能	MEMS/传感技术	发烧友学院	小组	PCB打板	新闻动态	hezuo@elecfans.com	lanhu@h
vr ar 虚拟现实	测量仪表	资料下载	博客	SMT加工	媒体报道	Mike Zhang	
可穿戴设备	嵌入式技术	技术专题	华秋众筹	元器件商城	加入我们	mikezhang@elecfans.com	投资合作
音视频及家电	制造/封装	39°	评测试用	定制开发	联系我们		曾海银
LEDs	模拟技术	Datasheet	VIP会员		侵权投诉	内容合作	zenghaiy
汽车电子	连接器	电路图		媒体服务		陆楠	
通信网络	EMC/EMI设计	电子百科		网站广告	社交网络	lunan@elecfans.com	原厂入驻
医疗电子	光电显示	元器件知识		在线研讨会	微博		杜女士：
智能电网	存储技术	在线工具		活动策划		内容合作（海外）	
安全设备/系统	EDA/IC设计	常用软件		新闻发布	移动	Brandon Smith	
军用/航空电子	处理器/DSP	电子书		新品发布	电子发烧友 APP	brandon.smith@elecfans.com	
移动通信	接口/总线/驱动			资源中心	电子发烧友 WAP	关注我们的微信	下载发烧
便携设备	控制/MCU			小测验			
触控感测	RF/无线			设计大赛			
工业控制				eDM			

[关于本站](#) | [欢迎投稿](#) | [用户建议](#) | [版权申明](#) | [友情链接](#) | [联系我们](#) | [网站地图](#)

版权所有 © 深圳华秋电子有限公司 电子发烧友粤ICP备14022951号

站长推荐  4/6 

关闭

**【开发板活动第1期】电子发烧友“开发板交换站”活动上线，一起来换积分！换板！换礼...**

本次活动为各位热忠开发板研究设计的发烧友量身打造，用你的闲置开发板来兑换成积分并可置换你想要的开发板或礼品！

[查看 »](#)