


通俗理解决策树算法中的信息增益



忆臻
哈尔滨工业大学 计算机科学与技术博士在读

关注他

185 人赞同了该文章

在决策树算法的学习过程中，信息增益是特征选择的一个重要指标，它定义为一个特征能够为分类系统带来多少信息，带来的信息越多，说明该特征越重要，相应的信息增益也就越大。

概念

我们前面说了，信息熵是代表随机变量的复杂度（不确定度）通俗理解信息熵 - 知乎专栏，条件熵代表在某个条件下，随机变量的复杂度（不确定度）通俗理解条件熵 - 知乎专栏。

而我们的信息增益恰好是：信息熵-条件熵。

换句话说，信息增益代表了一个条件下，信息复杂度（不确定性）减少的程度。

那么我们现在也很好理解了，在决策树算法中，我们的关键就是每次选择一个特征，特征有多个，那么到底按照什么标准来选择哪一个特征。

这个问题就可以用信息增益来度量。如果选择一个特征后，信息增益最大（信息不确定性减少的程度最大），那么我们就选取这个特征。

例子

我们有如下数据：

帅？	性格好？	身高？	上进？	嫁与否
帅	不好	矮	不上进	不嫁
不帅	好	矮	上进	不嫁
帅	好	矮	上进	嫁
不帅	烂好	高	上进	嫁
帅	不好	矮	上进	不嫁
帅	不好	矮	上进	不嫁
不帅	好	高	不上进	嫁
不帅	好	中	上进	嫁
帅	烂好	中	上进	嫁
不帅	不好	高	上进	嫁
帅	好	矮	不上进	不嫁
帅	好	矮	不上进	不嫁

可以求得随机变量X（嫁与不嫁）的信息熵为：

嫁的个数为6个，占1/2，那么信息熵为-1/2log1/2-1/2log1/2 = -log1/2=0.301

现在假如我知道了一个男生的身高信息。

身高有三个可能的取值{矮，中，高}

矮包括{1,2,3,5,6,11,12}，嫁的个数为1个，不嫁的个数为6个

中包括{8,9}，嫁的个数为2个，不嫁的个数为0个

高包括{4,7,10}，嫁的个数为3个，不嫁的个数为0个

先回忆一下条件熵的公式如下：

$$H(Y|X) = \sum_{x \in X} p(x) H(Y|X = x)$$

我们先求出公式对应的：

H(Y|X = 矮) = -1/7log1/7-6/7log6/7=0.178

H(Y|X=中) = -1log1-0 = 0

H(Y|X=高) = -1log1-0=0

p(X = 矮) = 7/12,p(X =中) = 2/12,p(X=高) = 3/12

则可以得出条件熵为：

7/12*0.178+2/12*0+3/12*0 = 0.103

那么我们知道信息熵与条件熵相减就是我们的信息增益，为

0.301-0.103=0.198

所以我们可以得出我们在知道了身高这个信息之后，信息增益是0.198

结论

我们可以知道，本来如果我对一个男生什么都不知道的话，作为他的女朋友决定是否嫁给他的不确定性有0.301这么大。

当我们知道男朋友的身高信息后，不确定度减少了0.198.也就是说，身高这个特征对于我们广大女生同学来说，决定嫁不嫁给自己的男朋友是很重要的。

至少我们知道了身高特征后，我们原来没有底的心里（0.301）已经明朗一半多了，减少0.198了（大于原来的一半了）。

那么这就类似于非诚勿扰节目里面的桥段了，请问女嘉宾，你只能知道男生的一个特征，请问你想知道哪个特征。

假如其它特征我也全算了，信息增益是身高这个特征最大。那么我可以说，孟非哥哥，我知道男嘉宾的一个特征是身高特征。因为它在这些特征中，对于我挑夫君是最重要的，信息增益是最大的，知道了这个特征，嫁与不嫁的不确定度减少的是最多的。

哈哈，希望能对理解信息增益有所帮助。

编辑于 2017-04-27

「真诚赞赏，手留余香」


赞赏

6 人已赞赏



信息论 自然语言处理 决策树

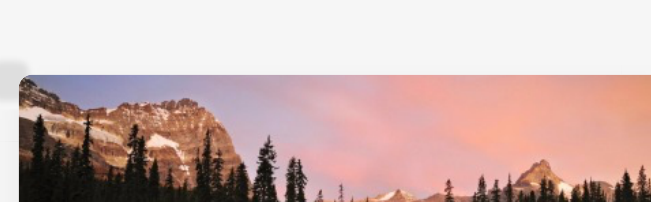
文章被以下专栏收录



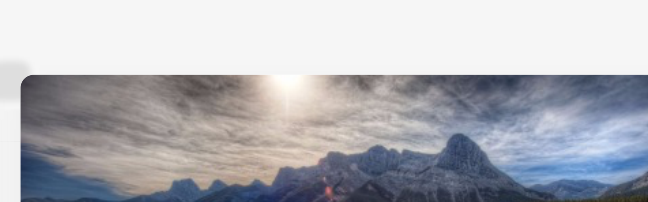
机器学习算法与自然语言处理
公号[机器学习算法与自然语言处理] 微信号yizhennotes

关注专栏

推荐阅读



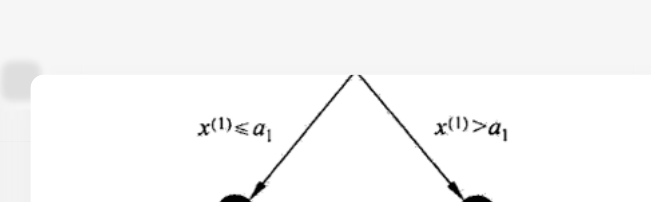
深入浅出理解决策树算法（二）-ID3算法与C4.5算法
忆臻 发表于机器学习算...



深入浅出理解决策树算法（一）-核心思想
忆臻 发表于机器学习算...

$$I(X) = - \sum_{i=1}^n p(x_i) \log p(x_i)$$

通俗理解信息熵
忆臻 发表于机器学习算...



决策树（一）熵、条件熵、信息增益
黑夜不再来

47 条评论

切换为时间排序

写下你的评论...

😊



深海

可以，我正在看这个

2017-04-27



忆臻 (作者) 回复 深海

应该对你有帮助

2017-04-27



深海 回复 忆臻 (作者)

是的是的，大神的资料效果还是好

2017-04-27

展开其他 3 条回复



何晔

看懂了。支持

2017-04-27



忆臻 (作者) 回复 何晔

谢谢~

2017-04-27



黑小贱

帮助很大。。付费一下

2017-05-16



忆臻 (作者) 回复 黑小贱

非常感谢

2017-05-16



诸葛冰箱

写得很好！！

2017-07-15



忆臻 (作者) 回复 诸葛冰箱

谢谢~

2017-07-15



乃乎

真的很感谢，看了这么多资料，西瓜书都没你这个讲的清楚了！！！！

2017-08-10



忆臻 (作者) 回复 乃乎

非常开心对你有帮助~

2017-08-10



eric 回复 忆臻 (作者)

大神，请教一下，看了你的文章中信息增益=信息熵-条件熵，进行多个特征相关性比较的时候，为什么不直接使用条件熵，因为条件熵大的，增益肯定小，直接选择条件熵小的不就可以了么？不太理解，请赐教

2017-10-10

查看全部 7 条回复



乐子

棒棒哒

2017-10-13



孔凡人

大神，我想问一下，为什么条件熵可以代表不包含特征词时的系统信息量

2017-11-04



Titans

太感谢了，讲的很清楚

2017-11-27



sluk3r

非诚勿扰中的桥段太形象了，赞👍

2017-11-28



jiazhuanh

很通俗易懂！

2018-01-09



johnKL

难道没有人觉得log应该是以2为底吗？能得出这个答案“-log1/2=0.301”，显然是以10为底

2018-01-23



忆臻 (作者) 回复 johnKL

这个不是重点，以谬为底

2018-01-23



Jason Chou 回复 johnKL

以2或10或e为底，我认为都是一样的，都是递增的函数

2019-02-27



茜子西

通俗易懂，感谢！

2018-01-25



ZORO LUFFY

感谢这个回答拯救了我的作业

2018-04-13



Eterian

谢谢哈，虽然还是有点模糊，不过例子很形象了。

2018-08-22



屈金强

这解释浅显易懂，牛逼的厉害

2018-09-13



战斧式巡航导弹

我要把你的文章全看完

2018-10-10



JokerZeroGxxx

关于女生找男朋友的那个，要选信息增益最大的，别只算身高呀，还有性格，给给男同胞点希望啊，难过.jpg

2018-10-31



傲来国主 回复 JokerZeroGxxx

主要还是钱的问题

2018-12-20



Thatkiddo

难为分类系统带来多少信息，带来的信息越多，说明该特征越重要（这句不太理解，信息量和概率不是成反比吗，请问给系统带来的信息是指剔除这个属性后剩下的信息量吗）

2019-01-01



王尊一

通俗易懂！！

2019-03-07



穆华龄

写的超棒，穷学生赞赏5元以表敬意。

2019-03-19