

# CART剪枝详解

1513

原创 MrTriste 最后发布于2017-08-06 20:20:25 阅读数 7497 ☆ 收藏

## CART之剪枝详解

我们这里用的是代价复杂度剪枝算法。

首先我们将一颗充分生长的树称为 $T_0$ ，我们希望减少树的大小来防止过拟合，但又担心去掉一些节点后预测的误差会增大，那: 达到这两个变量: 是问题的关键，因此我们用一个变量 $\alpha$  来平衡，因此损失函数定义为如下：

$$C_{\alpha}(T) = C(T) + \alpha|T|$$

T为任意子树，C(T)为预测误差，可以是平方误差也可以是基尼指数，|T|为子树T的叶子节点个数，注意是叶子节点， $\alpha$  是参数，C(T)衡量训练数据拟合程度，|T|衡量树的小)， $\alpha$  权衡拟合程度与树的复杂度。

那么我们如何找到这个合适的 $\alpha$ 来使拟合程度与复杂度之间达到最好的平衡呢，最好的办法就是，我们将 $\alpha$ 从0取到正无穷，对于每一个固定的 $\alpha$ ，我们使得 $C_{\alpha}(T)$ 最小的最优子树 $T(\alpha)$ 。当 $\alpha$  很小的时候， $T_0$  是这样的最优子树，当 $\alpha$  很大的时候，单独一个根节点是这样的最优的子树。

尽管 $\alpha$  取值无限多，但是 $T_0$  的子树是有限个，因此我们可以生成这样一个子树序列

$$T_0 > T_1 > T_2 > \dots > T_n$$

$T_n$  是最后剩下的那个根节点。（这里的子树生成是根据前一个子树 $T_i$ ，剪掉某一个内部节点，生成 $T_{i+1}$ ）然后对这样的子树序列分别用测试集进行交叉验证，找到最优的我们的决策树。

**高能预警：(这里不注意的话，后面会有很大的疑惑)**

**Breiman证明：将 $\alpha$  从小增大， $0 = \alpha_0 < \alpha_0 < \dots < \alpha_n < +\infty$ ，在每个区间 $[\alpha_i, \alpha_{i+1})$  中，子树 $T_i$  是这个区间里最优的。**

这也是代价复杂度剪枝的核心思想。

基于上面的论述，剪枝可分为两部分，第一部分生成子树序列，第二部分交叉验证。

### 1. 生成子树序列

我们每次剪枝剪的都是某个内部节点的子节点，也就是将某个内部节点的所有子节点回退到这个内部节点里，并将这个内部节点作为叶子节点。因此在失函数时，这个内部节点以外的值都没变，只有这个内部节点的局部损失函数改变了，因此我们本需要计算全局的损失函数，但现在只需要计算内部节点剪枝后的损失函数。

对任意内部节点t，

剪枝前的状态：有 $|T_t|$  个叶子节点，预测误差是 $C(T_t)$

剪枝后的状态：只有本身一个叶子节点，预测误差是 $C(t)$

因此剪枝前的以t节点为根节点的子树的损失函数是

$$C_{\alpha}(T_t) = C(T_t) + \alpha|T_t|$$

剪枝后的损失函数是

$$C_{\alpha}(t) = C(t) + \alpha$$

易得，一定存在一个 $\alpha$ 使得 $C_{\alpha}(T_t) = C_{\alpha}(t)$ ，这个值为

$$\alpha = \frac{C(t) - C(T_t)}{|T_t| - 1}$$

这个 $\alpha$  的值有什么意义，刚才我们高能预警的地方， $0 = \alpha_0 < \alpha_1 < \dots < \alpha_n < +\infty$ ，在每个区间 $[\alpha_i, \alpha_{i+1})$  中，子树 $T_i$  是这个区间里最优的，为什么呢？原因就在于当前这个节点，只要 $\alpha$  大于这个值时，一定有 $C_{\alpha}(t) < C_{\alpha}(T_t)$ ，也就是剪掉这个节点后都比不剪要更优。所以每个最优子树对应的是一个在这个区间内都是最优的

然后我们对 $T_i$  中的每个内部节点t都计算



举报

(注意这里的 $t$ 是变量，上面推 $\alpha$ 时是针对一个特定的 $t$ ，也就是 $t$ 是常量。)

书上说 $g(t)$ 表示剪枝后整体损失函数减少的程度，然后剪去 $g(t)$ 最小的 $T_t$ 。

当初这是非常困扰我的一个地方，1.为什么代表整体损失函数减少的程度。2.既然代表减少的程度，为什么剪去 $g(t)$ 最小的 $T_t$ ，去 $(C(t) + \alpha)$ 最小的 $T_t$ 。

- 1.首先针对第一个问题

我承认我到现在也没弄懂，我觉得写的有问题或者有歧义？我觉得把它解释成 **误差减小率** 更合适，我觉得它是衡量叶子节点增加一个节点带来的误差减小量，如果我们单一的看误差见效的量有可能会失偏颇，因为你可能是增加了很多很多的叶子节点才换来的这么一点误差减小量，因此我们用叶子节点数，来代表每个叶子节点带来的误差减小量。

- 2.为什么剪去 $g(t)$ 最小的 $T_t$

如刚才说的 $g(t)$ 代表每个叶子节点带来的误差减小量，如果 $g(t)$ 越小，也就是增加叶子带来的误差减小量越小，也就是增加这个叶子节点的作用越小，不去增加叶子节点，误差才减小那么一点点，还不如不要，因此优先剪去 $g(t)$ 最小的 $t$

- 3.为什么不剪去 $(C(t) + \alpha)$ 最小的 $T_t$ 。

考虑过这个问题的说明有一定思考了。对当前的 $T_t$ ，我们如果希望损失函数最小的话，应该是剪去损失函数最小的那个内部节点啊，这样的话损失函数最小嘛。

但是上面那句话真的对嘛？

整体损失函数 = 内部节点 $t$ 的损失函数 + 其他节点的损失函数和

如果我们剪去的损失函数最小的那个内部节点，那么等号右边第一项是最小了，那第二项呢？所以提出的疑问也就不成立了。

进一步，为什么会让人产生这样的疑问呢？我觉得主要是没弄明白节点 $t$ 是 **常量** 还是 **变量**。

**在计算 $\alpha$ 时，节点 $t$ 是常量，也就是只考虑这一个节点，当 $\alpha > g(t)$ 时，剪掉会使整体损失函数减小，但跟整体损失函数是否是最小没有关系。考虑 $\alpha$ 的因此**在计算整体的损失函数时，这个内部节点以外的值都没变，只有这个内部节点的局部损失函数改变了，因此我们本需要计算全局的损失函数，但现在只需要点剪枝前和剪枝后的损失函数。**这个只跟能否使整体损失函数减小有关，与整体损失函数能否达到最小无关。**

## 2. 交叉验证

用交叉验证法在子树序列中选取最优子树。

👍 点赞 15    ☆ 收藏    🔄 分享    ...



MrTriste

发布了35 篇原创文章 · 获赞 64 · 访问量 7万+

私信

亿速云香港服务器ping值低仅29元

买服务器选亿速云服务器,轻松上云,CN2高速连接,ping值低可免费换IP安全稳定,香港服务器售价29元

广告 亿速云



想对作者说点什么



A Han    2年前

查

整个过程归结来说就是用交叉验证尝试不同的阿尔法值。因为阿尔法是离散的，并且可以知道在每个不同的小区间上阿尔法值的作用各自都是一样的（因为在区间上的表的修剪出来的树是一样的），所以总结来说，上述的过程就是求解各个区间端点的过程。而端点就是不同的 $g(t)$ 。所以我觉得博主最后的三个问题，思考有点不太



1ingzero    9个月前

今天看到这篇文章觉得写得很好，关于“ $g(t)$ 表示剪枝后整体损失函数减少的程度，然后剪去 $g(t)$ 最小的 $T_t$ ”，我的理解是：剪枝的过程是将 $T_t$ 逐渐变得简单（对应 $\alpha$ 然后得到一系列决策树，最终通过验证集来确定最优决策树的，所以每次减去的是最小的



举报

登录 查看 13 条热评

15

15

13

13

13

13

13

13

13

13

13

13

13

13

13

13

13

13

13

13

13

13

13

13

13

13

13

13

13

13

13

13

13

13

13

13

## 决策树剪枝(cart剪枝)的原理介绍

最近看《统计学习方法》，最后有一部分讲到cart树的剪枝策略，个人觉得书上讲得比较晦涩难懂，现在结合个人理...

博文

来自: zhengzh

## cart决策树剪枝的个人理解

先说好前面部分我引用自以下这篇文章的，我不想写了https://blog.csdn.net/wjc1182511338/article/details/767...

博文

来自: wqtlm的

## CART剪枝算法

## 程序员实用工具网站

目录1、搜索引擎2、PPT3、图片操作4、文件共享5、应届生招聘6、程序员面试题库7、办公、开发软件8、高清图...

博文

来自: 不脱发的程

## 工业互联网平台，助力数字化转型

3月3日在线讲堂，分享技术和1+N生态快速帮助企业拥抱数字化时代的成功经验

广告 阿里云

## 统计学习方法第五章：决策树(decision tree),CART算法，剪枝及python实现

阅读数 1363

决策树(decision tree)是一种基本的分类与回归方法。决策树模型呈树状结构，在分类问题中，表示基于特征对实例...

博文

来自: devil\_bye的博客

## 决策树系列（五）——CART

阅读数 1715

原文地址CART，又名分类回归树，是在ID3的基础上进行优化的决策树，学习CART记住以下几个关键点：（1）CA...

博文

来自: Yan456jie的专栏

## 【机器学习】决策树（下）——CART算法及剪枝处理

阅读数 8491

CART，即分类与回归树（classification and regression tree），也是一种应用很广泛的决策树学习方法...

博文

来自: HerosOfEarth的博客

## CART剪枝算法详解

阅读数 1553

CART剪枝算法CART剪枝算法从“完全生长”的决策树的底端剪去一些子树，使决策树变小（模型变简单），从而能...

博文

来自: tian\_tian\_hero的...

## 机器学习--详解CART树剪枝原理和过程

阅读数 1277

这一节主要讲前面多次的提到的决策树问题，前面的决策树生成算法递归的产生决策树，直到不能继续分支或者...

博文

来自: 进击的菜鸟

## CART分类回归树

阅读数 7057

这篇主要是CART，有几个重点的词语先写下来，重点哦：基尼指数(Gini index，跟经济学的基尼系数长这么像，...

博文

来自: learner的专栏

## 后剪枝技术之代价复杂性剪枝

阅读数 4746

参考博文http://blog.csdn.net/tianguokaka/article/details/9018933在分类回归树中可以使用的后剪枝方法有多种...

博文

来自: zhang4418876的...



zhengzhenxian

4篇文章

关注 排名:千里之外



wqtlm

62篇文章

关注 排名:千里之外



zlsjsj

42篇文章

关注 排名:千里之外



不脱发的程序猿

156篇文章

关注 排名:944

## 数据挖掘之CART剪枝

阅读数 1192

与上篇文章中提到的ID3算法和C4.5算法类似，CART算法也是一种决策树分类算法。CART分类回归树算法的本质也...

博文

来自: cuiljuan02的博客

## 周志华《机器学习》课后习题解答系列（五）：Ch4.4 - 编程实现CART算法与剪枝操作

阅读数 2824

基于训练集构建的完全决策树易陷入过拟合，为提升模型泛化能力，通常需要对树进行剪枝。此处基于基尼系数构建...

博文

来自: Snoopy\_Yuan技术...

## 求决策树剪枝中的悲观剪枝pep详解，最好有例子说明。

求决策树剪枝中的悲观剪枝pep详解，最好有例子说明。

## CART树剪枝的操作的理解

这里我就简单讲下CART剪枝的核心思想，纯属个人意见，如有不当，请指正。在《统计学习方法》中已经提到了...

博文

来自: wjlwangluo的博客

论坛

举报

1054

@CART决策树剪枝个人理解本文是针对CART树的个人理解在看统计学习方法关于CART树的剪枝是，感觉书上讲得...

博文

来自: weixin\_3

50...

15

1万+

13

530

☆

40

90...

<

>

87

赏

阅读数 3万+

## 决策树 (Decision Tree) 进阶应用 CART剪枝方法及Python实现方式

决策树 Decision TreeC5.0先简述下C5.0，C5.0是一个商业软件，对于公众是不可得到的。它是在C4.5算法做了一些...

博文

来自: 新手村

## [机器学习 - 算法原理] CART树剪枝的理解

看了李航的《统计学习方法》中CART树的剪枝，不很清晰，又查了网上一些相关资源，终于明白了整个过程，个人...

博文

来自: maomaor

客

## CART后剪枝理解

转载https://blog.csdn.net/zhengzhenxian/article/details/79083643最近也是要找工作，在看李航的《统计学方...

博文

来自: weixin\_4

## Decision Tree: CART、剪枝

Decision Tree: CART、剪枝 决策树的重要性和入门可以参考前面两篇文章，尤其是入门的ID3算法: http://isilic.it...

博文

来自: Coding、

## 决策树之剪枝原理与CART算法

决策树学习笔记 (二) 继续关于决策树的内容，本篇文章主要学习了决策树的剪枝理论和基于二叉树的CART算法。...

博文

来自: Demon-初来驾到

## 决策树-Cart生成和剪枝算法

\*\*\*\*\*CART算法概述\*\*\*\*\* Cart算法类似于ID3算法，其将特征分类为按GINI系数，找到...

博文

来自: 每天进步一点点

阅读数 1835

## 决策树剪枝策略实例分析

预剪枝：是在决策树的生成过程中，对每个结点在划分前先进行估计，若当前结点的划分不能带来决策树泛化性能提...

博文

来自: toplatona

阅读数 2511

## 决策树原理及实现 (二) --CART算法及剪枝

本节介绍决策树中使用最多的CART方法以及剪枝，主要参考资料。在上一节我们将了ID3算法，和ID3算法的改进版...

博文

来自: crazy\_scott的博客

阅读数 725

## CART算法实现之剪枝篇

[Download Source code]

博文

来自: 我为编程狂

阅读数 3594

## 决策树\_剪枝

首先剪枝 (pruning) 的目的是为了避免决策树模型的过拟合。因为决策树算法在学习的过程中为了尽可能的正确的...

博文

来自: AcceptedLin的博客

阅读数 865

## 决策树CART以及剪枝

上一篇文章写了C4.5和ID3，这一篇介绍另外一种常用的决策树CART，CART是后续学习GBDT，Xgboost需要用到...

博文

来自: poppyabrown的...

阅读数 1087

## 阿里巴巴2020年校招开始啦

老师，您好！我是黄威,2011年武汉毕业的，现在在阿里巴巴从事软件开发，希望学弟学妹过来实习实习地点：杭州...

博文

来自: weixin\_34315189...

阅读数 449

## Alpha-beta剪枝算法实例分析

看本章之前，请先参看前一篇文章《Minimax算法及实例分析》由于Minimax算法有一个很大的问题就是计算复杂性...

博文

来自: 我的专栏

阅读数 1万+

## 我的阿里秋招之路——阿里实习offer+校招offer

我的2017是忙碌的一年，从年初备战实习春招，年三十都在死磕JDK源码，三月份经历了阿里五次面试，四月顺利收...

博文

来自: 大闲人柴毛毛

阅读数 4万+

## 李航，统计学习方法：CART剪枝算法实例分析

目录1.附上剪枝算法2.符号解析首先子树的概念再看公式：3.真实案例算法第三步的实例计算从人的角度观察：1.附...

博文

来自: qq\_30636613的博客

阅读数 85

## CART树算法的剪枝算法

输入是CART树建立算法得到的原始决策树T。输出是最优决策子树T $\alpha$ 。算法过程如下：1) 初始化 $\alpha_{min} = \infty$ ，最优子...

博文

来自: zrh\_CSDN的博客

阅读数 355

## 《统计学习方法》笔记——5.5.2CART剪枝

在看《统计学习方法》里的 5.5.2 CART剪枝 时，书中有一句话让我很迷惑——g(t)表示剪枝后整体损失函数减少的...

博文

来自: 章小幽的博客

阅读数 132

## cart决策树的后剪枝策略

后剪枝策略一般要求我们先构建好决策树，使得每个叶子节点都是纯的（只有一种结果）。然后再进行剪枝操作。...

博文

来自: weixin\_3

40...

举报

## 阿里校招四面经验分享

java,阿里，阿里巴巴，面试经验，经验分享，阿里面经

博文

来自: 狼图的点滴

阅读数 1936

分类回归树(CART,Classification And Regression Tree)也属于一种决策树，上回文我们介绍了基于ID3算法的决策...

博文 来自: weixin\_3 16...

## 决策树剪枝的方法与必要性

zhuan1 决策树剪枝的必要性本文讨论的决策树主要是基于ID3算法实现的离散决策树生成。ID3算法的基本思想是贪...

博文 来自: yeting067 533

## 决策树的剪枝理论

文章出处: [http://blog.sina.com.cn/s/blog\\_4e4dec6c0101fdz6.html](http://blog.sina.com.cn/s/blog_4e4dec6c0101fdz6.html)剪枝理论，决策树的剪枝在上一节中没有仔...

博文 来自: u0105601 524

## Opencv2.4.9源码分析——Decision Trees

一、原理 决策树是一种非参数的监督学习方法，它主要用于分类和回归。决策树的目的是构造一种模型，使之能够从...

博文 来自: 赵春江的 482

## 爬虫福利二 之 妹子图网MM批量下载

爬虫福利一: 27报网MM批量下载 点击看了本文，相信大家对爬虫一定会产生强烈的兴趣，激励自己去学习爬虫，...

博文 来自: Nick.Peng 万+

python json java mysql pycharm android linux json格式 c#账号密码验证 代码 c# 点击保存草稿的实现 c# 如何切  
界面 隐藏部分 c#关闭套接字 c#报表怎么绑定图片 c# id读写器 c#俄罗斯方块源码 c# linq原理 c# 装箱有什么用



MrTriste

TA的个人主页 >

原创 35 粉丝 46 获赞 64 评论 58 访问 7万+

等级: 博客 4 周排名: 11万+

积分: 1085 总排名: 7万+

勋章:

关注

私信



LOUIS VUITTON

### 最新文章

Kaggle比赛分类与winner资料汇总 (更新中)

restore与  
run(tf.global\_variables\_initializer())

tensorflow 恢复(restore)模型的两种方式

强化学习-An introduction之 On-policy  
Prediction with Function Approximation  
个人笔记

python3 str与bytes的问题

### 分类专栏



Machine Learning In...

8篇



举报



C++



深度学习

5篇



神经网络

2篇

展开

归档

2018年11月	1篇
2018年8月	2篇
2018年7月	7篇
2018年6月	1篇
2018年4月	8篇
2018年3月	1篇
2018年2月	1篇
2018年1月	1篇

展开

热门文章

LSTM详解 反向传播公式推导

阅读数 18117

TensorFlow实战：LSTM的结构与cell中的参数

阅读数 9215

循环神经网络RNN详解 反向传播公式推导+代码（十分详细）

阅读数 8334

CART剪枝详解

阅读数 7494

tensorflow 恢复(restore)模型的两种方式

阅读数 6080

最新评论

TensorFlow实战：LSTM...

hz\_jhx: [reply]wjc1182511338[/reply]您好，这个图链接https://ws1.sinaimg.cn/large/006g ...

CART剪枝详解

mywenw: 交叉验证选取最优子树？难道是用交叉验证集选一个最小的g(t)就好了？这样的话，仍 ...

CART剪枝详解

mywenw: 补充一句，对任意内部节点t，T\_t是以t为根节点的子树

CART剪枝详解

mywenw: [reply]appleychi[/reply] 你这里说的迭代是指什么的迭代呢？

TensorFlow实战：LSTM...

m0\_37834186: n\_step是20吗？



15



13



举报



LOUIS VUITTON

🗨️ QQ客服

✉️ kefu@csdn.net

🗣️ 客服论坛

☎️ 400-660-0108

工作时间 8:30-22:00

关于我们 招聘 广告服务 网站地图

京ICP备19004658号 经营性网站备案信息

👮 公安备案号 11010502030143

©1999-2020 北京创新乐知网络技术有限  
公司 网络110报警服务

北京互联网违法和不良信息举报中心

中国互联网举报中心 家长监护 版权申诉

👍

15

🔗

💬

13

☆

📱

<

>

赏

🔊

举报