

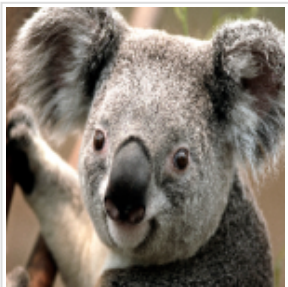
lujiandong1的专栏

目录视图

摘要视图

RSS 订阅

个人资料



BYR_jiandong



访问：132451次

积分：3525

等级：BLOG > 5

排名：第8044名

原创：220篇 转载：41篇

译文：0篇 评论：22条

【有奖投票】玩转Dragonboard 410c 的正确姿势 CSDN日报20170406 —— 《代码很烂，所以离职。》 Python数据分析与机器学习 博客搬家，有礼相送

pandas使用get_dummies进行one-hot编码

标签：pandas one-hot

2016-10-17 09:30

1724人阅读

评论(0)

收藏

举报

分类：机器学习 (32) ▼

版权声明：本文为博主原创文章，未经博主允许不得转载。

离散特征的编码分为两种情况：

- 1、离散特征的取值之间没有大小的意义，比如color：[red,blue],那么就使用one-hot编码。
- 2、离散特征的取值有大小的意义，比如size:[X,XL,XXL],那么就使用数值的映射。

使用pandas可以很方便的对离散型特征进行one-hot编码

关闭



文章搜索

文章分类

ACM DP问题 (7)

ACM其他文章 (2)

STL (5)

人生感悟 (1)

C++基础 (20)

C# (1)

ACM 贪心算法 (2)

ACM 哈希 (3)

二分法 (3)

STL 空间配置器 (1)

poj Trie树 (1)

poj 调试经验 (2)

POJ KMP (2)

poj 暴力法 (1)

C++调试 (3)

More Effective C++读书笔记 (3)

Effective C++读书笔记 (25)

机器学习 (33)

北邮人论坛 (2)

杂项 (6)

数学 (2)

opencv (3)

操作系统基本概念 (10)

Linux基础学习 (10)

程序优化 (3)

深入理解计算机系统 (5)

[python]  

```

01. import pandas as pd
02. df = pd.DataFrame([
03.     ['green', 'M', 10.1, 'class1'],
04.     ['red', 'L', 13.5, 'class2'],
05.     ['blue', 'XL', 15.3, 'class1']])
06.
07. df.columns = ['color', 'size', 'prize', 'class label']
08.
09. size_mapping = {
10.     'XL': 3,
11.     'L': 2,
12.     'M': 1}
13. df['size'] = df['size'].map(size_mapping)
14.
15. class_mapping = {label:idx for idx,label in enumerate(set(df['class label']))}
16. df['class label'] = df['class label'].map(class_mapping)

```

说明：对于有大小意义的离散特征，直接使用映射就可以了，{'XL':3,'L':2,'M':1}

	color	size	prize	class label
0	green	1	10.1	0
1	red	2	13.5	1
2	blue	3	15.3	0

Using the get_dummies will create a new column for every unique string in a
行one-hot编码

[python]  

```

01. pd.get_dummies(df)

```

关闭



[Python 基础](#) (25)[SVM](#) (1)[推荐系统](#) (1)[机器学习讲座笔记](#) (1)[Kaggle学习笔记](#) (10)[特征工程](#) (2)[caffe教程及遇到问题的解决方案](#) (2)[leetcode](#) (7)[自然语言处理](#) (10)[哈工大SCIR 神经网络和深度学习转载](#) (11)[tensorflow调研](#) (37)

文章存档

[2017年03月](#) (5)[2017年02月](#) (9)[2017年01月](#) (2)[2016年12月](#) (8)[2016年11月](#) (31)[展开](#)

阅读排行

[安装scikit-learn , win7 64](#) (8805)[SVM的两个参数 C 和 gamma](#) (6803)[Python 列表的清空](#) (3361)[import sys sys.path.append](#) (3062)[error LNK2019: 无法解析](#) (3018)[连续特征离散化达到更好](#) (2567)[配置caffe的python接口及](#) (2194)

	size	prize	class label	color_blue	color_green	color_red
0	1	10.1	0	0	1	0
1	2	13.5	1	0	0	1
2	3	15.3	0	1	0	0

顶

0

踩

0

[上一篇](#) [xgboost cross_validation&自定义目标函数和评价函数&base_score参数](#)[下一篇](#) [安装jpype出现_jpype错误的解决方案](#)

我的同类文章

机器学习 (32)

- [LSTM的完整推导过程，附上..](#) 2016-11-16 阅读 239
- [机器学习常见的六大错误](#) 2016-11-02 阅读 111
- [xgboost cross_validation&自...](#) 2016-10-11 阅读 896
- [AUC的理解与应用场景](#) 2016-07-23 阅读 167
- [随机森林的几个重要](#)
- [安装jpype出现_jpype错误的解决方案](#)
- [XGBoost Stopping t](#)
- [k-means优化 & k-m](#)

[关闭](#)

[caffe 教程 Fine-tuning C](#) (1938)[机器学习中的内核方法](#) (1873)[DBN的训练过程](#) (1790)

评论排行

[连续特征离散化达到更好](#) (4)[machine learning week6](#) (3)[error LNK2019: 无法解析](#) (2)[深入分析C++引用](#) (1)[人为什么会浮躁](#) (1)[Python 列表的清空](#) (1)[XGBoost Stopping to Av](#) (1)[Andrew Ng的 Machine L](#) (1)[配置caffe的python接口及](#) (1)[caffe 教程 Fine-tuning C](#) (1)

推荐文章

[* Android安全防护之旅---带你把Apk混淆成中文语言代码](#)[* TensorFlow文本摘要生成 - 基于注意力的序列到序列模型](#)[* 创建后台任务的两种代码模式](#)[* 一个屌丝程序员的人生（六十）](#)[* WKWebView与js交互之完美解决方案](#)[* 年轻人，“砖砖瓦瓦”不应该成为你的梦想！](#)[vowpal_wabbit是在单机上性...](#) 2015-12-30 阅读 514[机器学习经典博客链接](#)

2015-06-15 阅读 409

[更多文章](#)新东方
XDF.CN

北京总部

美国夏令营
探访世界名校 体验留学生活

马上报名

猜你在找

[ArcGIS for javascript 项目实战（环境监测系统）](#)[ArcGIS for JavaScript](#)[C# For Unity系列之基础篇](#)[ComponentOne Studio for WinForms基础课程](#)[C# For Unity系列之入门篇](#)[get方法带有汉字传递时需要对汉字进行编码如搜索操作](#)[Networking All-in-One Desk Reference For Dummies](#)[Dreamweaver 8 All-in-One Desk Reference For Dummies](#)[Macromedia Studio 8 All-in-One Desk Reference For](#)[Outlook 2007 All-in-One Desk Reference For Dummies](#)新东方
XDF.CN

北京

欧洲夏令营
欧洲文化探索
体验留学生活

了解详情



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

核心技术类目

[全部主题](#)[Hadoop](#)[AWS](#)[移动游戏](#)[Java](#)[Android](#)[iOS](#)[Swift](#)[管](#)[VPN](#)[Spark](#)[ERP](#)[IE10](#)[Eclipse](#)[CRM](#)[JavaScript](#)[数据库](#)[Ubu](#)

百度云

云计算新用户
注册送520元代金券

立即领取

关闭

最新评论

Andrew Ng的 Machine Learning fupf1303: 写的挺好的,可惜只有2和4,有其他的课程笔记吗?

tensorflow MNIST数据集上简单: 倾城一少: 博主,MLP网络的全称是什么?

tensorflow CNN for mnist xjbada: 我运行这个代码为什么会出现这个错误呢*** TypeError: __init__() got an...

tensorflow中关于队列使用的实验 yuehanliushuang: very good

tesnsorflow 使用LSTM进行分类 qq_27590277: 为什么说我出错 TypeError: __init__() got an unexpected ...

error LNK2019: 无法解析的外部: m0_37640107: 多谢博主! 我的opencv用vs生成的时候也一直报这个错误,上网其他方案都没用,直到看了你的才发现...

machine learning week6 诊断机: Starry5cm: rand_seq=round(rand(1,i)*(m-1))+1;%生成i个随机序列 0~m这里改...

machine learning week6 诊断机:

在国家 中国内地(China)

机号码

校验码

用户名

密码

确认密码

短信验证码接口



BI HTML5 Spring Apache .NET API HTML SDK IIS Fedora XML LBS Unity
 Splashtop UML components Windows Mobile Rails QEMU KDE Cassandra CloudStack FTC
 coremail OPhone CouchBase 云计算 iOS6 Rackspace Web App SpringSide Maemo
 Compuware 大数据 aptech Perl Tornado Ruby Hibernate ThinkPHP HBase Pure Solr
 Angular Cloud Foundry Redis Scala Django Bootstrap

关闭



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 400-600-2320 | 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 | 江苏乐知网络技术有限公司
京 ICP 证 09002463 号 | Copyright © 1999-2016, CSDN.NET, All Rights Reserved



百度云

云计算新用户
注册送520元代金券

立即领取