

[博客](#) (<http://blog.csdn.net/>)
[学院](#) (<http://edu.csdn.net/>)
[下载](#) (<http://download.csdn.net/>)
[GitChat](#) (<http://gitbook.cn/?ref=csdn>) [论坛](#) (<http://bbs.csdn.net/>)
...

写博客

发Chat

登录 (<https://passport.csdn.net/account/login>)

注册 (<https://passport.csdn.net/account/mobileregister?action=mobileRegister>)

pandas将类别属性转化为数值属性的方法

0
转载

2017年07月19日 22:48:00

1261

原文地址 (<http://blog.csdn.net/lujiandong1/article/details/52836051>)

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{X:1,XL:2,XXL:3}

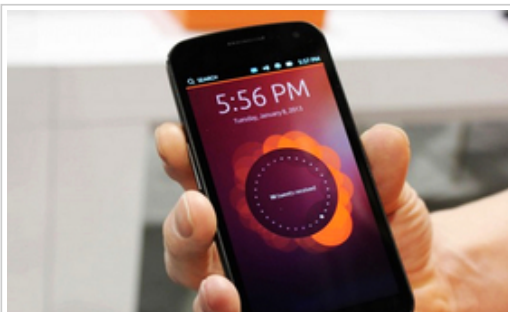
使用pandas可以很方便的对离散型特征进行one-hot编码



chenpe32cp (<http://blog.csdn.net/chenpe32cp>)

+ 关注

原创	粉丝	喜欢	未开通
12	5	2	(https://github.com/chenpe32cp)



ubuntu



他的最新文章

更多文章 [登录](#) (<https://passport.csdn.net/account/login>)



加入CSDN，享受更精准的内容推荐，与500万程序员共同成长！

```

1 import pandas as pd
2 df = pd.DataFrame([
3     ['green', 'M', 10.1, 'class1'],
4     ['red', 'L', 13.5, 'class2'],
5     ['blue', 'XL', 15.3, 'class1']])
6
7 df.columns = ['color', 'size', 'prize', 'class label']
8
9 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 certain column: 使用 get_dummies进行one-hot 编码

加入CSDN，享受更精准的内容推荐，与500万程序员共同成长！

如何判断DataFrame中是否有缺失值? <http://blog.csdn.net/chenpe32cp/article/details/78398090>

基于情感字典的文本情感极性分析 (<http://blog.csdn.net/chenpe32cp/article/details/77801600>)

对DataFrame列名进行更改的两种方式 (<http://blog.csdn.net/chenpe32cp/article/details/76408512>)

python3中无法导入cPickle (<http://blog.csdn.net/chenpe32cp/article/details/75212521>)

如何入门深度学习 (<http://blog.csdn.net/chenpe32cp/article/details/73188384>)

文章分类

python (<http://blog.csdn.net/c...>) 15篇
 java (<http://blog.csdn.net/chen...>) 1篇
 自然语言处理 (<http://blog.csdn.net/...>) 2篇
 mysql (<http://blog.csdn.net/ch...>) 1篇
 数据分析 (<http://blog.csdn.net/...>) 2篇

展开 ∨

文章存档

2017年10月 (<http://blog.csdn.net/...>) 1篇

	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



0



pandas 数据规整



liujianfei526

2016年01月05日 19:08

6426

合并数据集使用键参数的 DataFrame 合并轴向连接合并重叠数据重塑和轴向旋转重塑层次化索引将“长格式”转换为“宽格式”数据转换移除重复数据利用函数或映射进行数据转换替换值重命名轴索引离散化...

(<http://blog.csdn.net/liujianfei526/article/details/50464614>)

python dataframe astype 字段类型转换



chinacmt

2016年08月17日 13:59

34483

使用astype实现dataframe字段类型转换# -*- coding: UTF-8 -*- import pandas as pd df = pd.DataFrame({'col1':'a',...

(<http://blog.csdn.net/chinacmt/article/details/52230339>)

加入CSDN，享受更精准的内容推荐，与500万程序员共同成长！

2017年9月 (<http://blog.csdn.net/chenpe32cp/article/details/75452592>) 1篇

2017年7月 (<http://blog.csdn.net/chenpe32cp/article/details/75452592>) 3篇

2017年6月 (<http://blog.csdn.net/chenpe32cp/article/details/75452592>) 1篇

2017年5月 (<http://blog.csdn.net/chenpe32cp/article/details/75452592>) 1篇

展开 ∨

他的热门文章

使用python + selenium爬取淘宝商品信息
(<http://blog.csdn.net/chenpe32cp/article/details/54232353>)

3371

pandas将类别属性转化为数值属性的方法
(<http://blog.csdn.net/chenpe32cp/article/details/75452592>)

1244

使用python爬取《长城》豆瓣影评 (<http://blog.csdn.net/chenpe32cp/article/details/54180453>)

966

使用requests爬取豆瓣《长城》影评 (<http://blog.csdn.net/chenpe32cp/article/details/54232182>)

613

使用python_selenium爬取动态网页时要注意的问题 (<http://blog.csdn.net/chenpe32cp/article/details/54170878>)

476

登录

注册



一个数学公式教你秒懂天下英语



老司机教你一个数学公式秒懂天下英语

python+pandas+时间、日期以及时间序列处理

python、datetime, datetime.timedelta



LY_ysys629 2017年06月27日 19:36 22445

(http://blog.csdn.net/LY_ysys629/article/details/73822716)

十分钟搞定pandas (一)



qq826364410 2017年10月10日 11:07 142

这是关于pandas的简短介绍，主要面向新用户。可以参阅Cookbook了解更复杂的使用方法。习惯上,我们做以下导入 ...

(<http://blog.csdn.net/qq826364410/article/details/78190450>)

python中的字符数字之间的转换函数



liuweiyuxiang 2017年05月01日 21:13 344

int(x [,base]) 将x转换为一个整数 long(x [,base]) 将x转换为一个长整数 float(x) ...

(<http://blog.csdn.net/liuweiyuxiang/article/details/71075365>)

不再死记硬背，一个公式学懂英文



英语长难句解读，记住这个公式就够了！

python pandas将值转换为属性 / 特征



yepeng2007fei 2017年09月08日 15:54 187

问题：在进行数据分析的过程中我们常常需要将值转换为属性或者特征，例如： 转成 我写了段python代码做这方面
加入CSDN#享受更精准的内容推荐，与500万程序员共同成长！

(<https://passport.csdn.net/a>)

云虚拟主机-免费版

每人限购一台，续费后50天，永久免费。续费期间不限时，续费期间不限时。

1G	10G	50M
硬盘空间	月流量	数据量

立即购买 续费升级 续费升级 续费升级

便宜云虚拟主机

立即购买 续费升级 续费升级 续费升级

联系我们

网站客服 (<http://wpa.qq.com/msgrd?v=3&uin=2431299880&site=qq&menu>)

webmaster@csdn.net
(mailto:webmaster@csdn.net)

微博客服
(<http://e.weibo.com/csdnsupport/profile>)
400-660-0108

关于 招聘 广告服务

©2018 CSDN 京ICP证09002463号

(<http://www.miibeian.gov.cn/>)

经营性网站备案信息

(<http://www.hd315.gov.cn/beian/view.asp?>

bianhao=010202001032100010) 网络110

报警服务 (<http://www.cyberpolice.cn/>)

登录

注册



(<http://blog.csdn.net/yepeng2007fei/article/details/77895764>)

(<https://passport.csdn.net/a>)

pandas 下的重塑数据函数pivot- (转)



toyijiu 2018年01月17日 15:36 10

在CSV文件当中或者数据库当的数据通常以长格式或者 (stacked)堆叠格式存储，特别是金融数据中出现的时间序列数据，例如：In [1]: df Out[1]: date va...

(<http://blog.csdn.net/toyijiu/article/details/79086015>)



0

pandas的基本操作



sinat_29957455 2018年01月03日 22:44 39

1、reindex重新索引 pandas提供了一个reindex方法来创建一个适应新索引的新对象，Series通过调用reindex方法会根据新索引的顺序重新排序，如果新的索引中存在原索引中不存在...



(http://blog.csdn.net/sinat_29957455/article/details/78966383)

python—pandas中DataFrame类型数据操作函数

python数据分析工具pandas中DataFrame和Series常用函数简介.



LY_ysys629 2017年01月16日 11:29 13422

(http://blog.csdn.net/LY_ysys629/article/details/54428838)

Python 数据分析：pandas 操作基础篇



Eddy_zheng 2016年03月17日 11:14 6475

基础介绍pandas 是基于 Numpy 构建的含有更高级数据结构和工具的数据分析包类似于 Numpy 的核心是 ndarray，pandas 也是围绕着 Series 和 DataFrame 两个核...

(http://blog.csdn.net/Eddy_zheng/article/details/50912079)

pandas利用函数或映射进行数据转换



cuihuijun1hao 2017年10月16日 14:04 194

data = DataFrame({'food':['bacon','pull_pork','bacon','Pastrami','corned beef','Bacon','pastrami','h...

(<http://blog.csdn.net/cuihuijun1hao/article/details/78249064>)

加入CSDN，享受更精准的内容推荐，与500万程序员共同成长！

登录

注册



Pandas 数据处理（下）



u011532367

2016年03月30日 11:30

3848

df.apply(lambda x: x.max() - x.min()) 输出：A 2.073961 B 2.671590
C 1.785291 D 0.000000 F ...

(<http://blog.csdn.net/u011532367/article/details/51012902>)

(<https://passport.csdn.net/a>)

Python Pandas找到缺失值的位置



u012387178

2016年09月18日 09:44

17785

问题描述：python pandas判断缺失值一般采用 isnull()，然而生成的却是所有数据的true / false矩阵，对于庞大的数据dataframe，很难一眼看出来哪个数据缺...

(<http://blog.csdn.net/u012387178/article/details/52571725>)

Pandas数据特征分析



zengxiantao1994

2017年07月17日 10:32

1407

Pandas数据特征分析 数据的排序 .sort_index()方法在指定轴上根据索引进行排序，默认升序。默认0轴升序：.sort_index(axis=0, ascending=True)。 ...

(<http://blog.csdn.net/zengxiantao1994/article/details/75219388>)

Python如何将百分号的字符转成数字



qq_28219759

2016年10月31日 12:10

3456

原文链接：<http://www.datastudy.cc/to/63> 很多时候，我们拿到的报表数据，里面的小数都是带百分号的字符串，不能进行直接的计算，需要对其进行转换，然后再输出。 解决方...

(http://blog.csdn.net/qq_28219759/article/details/52981336)

基于pandas进行数据预处理



u014400239

2017年04月27日 11:10

1233

参加kaggle数据挖掘比赛，就第一个赛题Titanic的数据，学习相关数据预处理以及模型建立，本博客关注基于pandas进行数据预处理过程。包括数据统计、数据离散化、数据关联性分析引入包和加载数据i...

(<http://blog.csdn.net/u014400239/article/details/70846634>)

加入CSDN，享受更精准的内容推荐，与500万程序员共同成长！

登录

注册



几个pandas数据处理中的常用操作



zhaohansk

2016年01月08日 09:38

📖 15826

(https://passport.csdn.net/a

记几个常用的函数，具体用法去官网查啦 import的惯例：import pandas as pd from pandas import Series,DataFrame 显示一些数字特征：df.d...

(http://blog.csdn.net/zhaohansk/article/details/50480408)

数据对象的属性（特征）类型



zxd1754771465

2017年07月19日 09:43

📖 207

在数据挖掘/机器学习过程中，当我们在对属性进行处理时，分清属性的类型是有必要的，不同类型的属性需要使用不同的处理方法，如在对对象的标识列进行处理时，对该列进行数学运算是没有意义的，因为对象的标识列一般...

(http://blog.csdn.net/zxd1754771465/article/details/75334443)



python pandas中series与dataframe数据类型属性及操作基础

python编程pandas工具包中series和dataframe数据类型的属性及创建方法



LY_ysys629

2017年02月09日 14:30

📖 19452

(http://blog.csdn.net/LY_ysys629/article/details/54944153)

pandas或者sklearn中如何将字符形式的标签数字化

pandas中是很容易实现的,转换成Categories对象即可。术语叫做因子和水平，水平一般都会自动转成数字储存。c =

['A','A','A','B','B','C','C','C','C']...

(http://blog.csdn.net/Datapad/article/details/77852130)

加入CSDN，享受更精准的内容推荐，与500万程序员共同成长！

登录

注册

