

[CSDN首页 \(http://www.csdn.net?ref=toolbar\)](http://www.csdn.net?ref=toolbar)[学院 \(http://edu.csdn.net?ref=toolbar\)](http://edu.csdn.net?ref=toolbar)[下载 \(http://download.csdn.net?ref=toolbar\)](http://download.csdn.net?ref=toolbar)

更多 ▼

[下载 CSDN APP \(http://www.csdn.net/app?ref=toolbar\)](http://www.csdn.net/app?ref=toolbar)[写博客 \(http://write.blog.csdn.net/postedit?ref=toolbar\)](http://write.blog.csdn.net/postedit?ref=toolbar)[登录 \(https://passport.csdn.net/account/login?ref=toolbar\)](http://blog.csdn.net/) | [注册 \(http://passport.csdn.net/account/mobile/register?ref=toolbar&action=mobileRegister\)](http://passport.csdn.net/account/mobile/register?ref=toolbar&action=mobileRegister)[首页 \(http://blog.csdn.net/\)](http://blog.csdn.net/) [移动开发 \(http://blog.csdn.net/mobile/newarticle.html\)](http://blog.csdn.net/mobile/newarticle.html) 全部 □**CSDN** (<http://www.csdn.net>)

喜欢



收藏



评论



分享

## tensorflow中tfrecords文件的save和read

原创 2017年06月01日 13:14:01

328

0

0

在tensorflow程序中，推荐使用tensorflow内定标准格式——TFRecords，这是一种通用的有利于高效读取文件。TFRecords其实是一种二进制文件，虽然它不如其他格式好理解，但是它能更好的利用内存，更方便复制和移动，并且不需要单独的标签文件。

TFRecords文件包含了tf.train.Example 协议内存块(protocol buffer)(协议内存块包含了字段 Features)。我们可以写一段代码获取你的数据，将数据填入到Example协议内存块(protocol buffer)，将协议内存块序列化为一个字符串，并且通过tf.python\_io.TFRecordWriter 写入到TFRecords文件。

从TFRecords文件中读取数据，可以使用tf.TFRecordReader的tf.parse\_single\_example解析器。这个操作可以将Example协议内存块(protocol buffer)解析为张量。

下面我们直接通过代码片段体会TFRecords的生成和读取显示

HaiwiSong (<http://blog.cs...>)

+ 关注

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

码云

原创

粉丝

喜欢

未开通

57

6

0

(<https://github.com/oTengYue>)

### 他的最新文章

更多文章 (<http://blog.csdn.net/oTengYue>)

CRF模型在NLP中的运用

(/otengyue/article/details/78005557)

tensorflow中tfrecords文件的save和read

(/otengyue/article/details/72830262)

tensorflow设置log等级

(/otengyue/article/details/72823285)

hive中UDF开发：解析json对象和解析json数组对象

(/otengyue/article/details/70255202)



返回顶部



```

1  # -*- coding: utf-8 -*-
2  import argparse
3  import sys
4  import pandas as pd
5  import tensorflow as tf
6  import os
7  os.environ['TF_CPP_MIN_LOG_LEVEL']='3'
8
9  FLAGS = None
10  """
11  创建并生成tfrecords文件
12  """
13  def saveTfRecords(data_set, name):
14      user_id = data_set.user_id
15      age = data_set.age
16      sex = data_set.sex
17      user_lv_cd = data_set.user_lv_cd
18      user_reg_dt = data_set.user_reg_dt
19
20      filename = os.path.join(FLAGS.dir_path, name + '.tfrecords')
21      writer = tf.python_io.TFRecordWriter(filename)
22      for index in range(user_id.size):
23          # print(age[index])
24          example = tf.train.Example(features=tf.train.Features(feature={
25              'user_id': tf.train.Feature(int64_list = tf.train.Int64List(value=[user_id[index]])),
26              'age': tf.train.Feature(bytes_list = tf.train.BytesList(value=[str.encode(str(age[index]))])),
27              'sex': tf.train.Feature(float_list = tf.train.FloatList(value=[sex[index]])),
28              'user_lv_cd': tf.train.Feature(int64_list = tf.train.Int64List(value=[user_lv_cd[index]])),
29              'user_reg_dt': tf.train.Feature(bytes_list = tf.train.BytesList(value=[str.encode(str(user_reg_dt[index]))]))
30          }))
31          writer.write(example.SerializeToString())
32      writer.close()
33  """
34  读取tfrecords文件
35  """
36  def readTfRecords(name):
37      filename = os.path.join(FLAGS.dir_path, name + '.tfrecords')

```

编辑推荐

最热专栏

Notes on tensorflow (八) read tfrecor...

Tensorflow读取数据2-tfrecord (/u01091...

Tensorflow高阶读写教程 (/u010223750...

7.2 TensorFlow笔记(基础篇): 生成TFR...

## 在线课程



【免费】深入理解Docker  
([http://edu.csdn.net/course/detail/563?](http://edu.csdn.net/course/detail/563?utm_source=blog9)  
内部原理及网络配置  
(<http://edu.csdn.net/huiyi>  
Course/detail/563?  
blog9)



SDCC 2017之区块链技术应用  
([http://edu.csdn.net/course/series\\_detail/66?](http://edu.csdn.net/course/series_detail/66?utm_source=blog9)  
实战线上峰会  
(<http://edu.csdn.net/huiyi>  
Course/series\_detail/66?  
utm\_source=blog9)



返回顶部



喜欢



收藏



评论



分享

```

38 filename_queue = tf.train.string_input_producer([filename])
39 reader = tf.TFRecordReader()
40 _,serialized_example = reader.read(filename_queue)
41 features = tf.parse_single_example(
42     serialized_example,
43     features={
44         'user_id': tf.FixedLenFeature([],tf.int64),
45         'age': tf.FixedLenFeature([],tf.string),
46         'sex': tf.FixedLenFeature([],tf.float32),
47         'user_lv_cd': tf.FixedLenFeature([],tf.int64),
48         'user_reg_dt': tf.FixedLenFeature([],tf.string),
49     })
50 user_id = features['user_id']
51 age = features['age']
52 sex = features['sex']
53 user_lv_cd = features['user_lv_cd']
54 user_reg_dt = features['user_reg_dt']
55 return user_id,age,sex,user_lv_cd,user_reg_dt
56
57 """
58 读取csv文件
59 """
60 def getDataSet(file_path):
61     csv = pd.read_csv(file_path)
62     return csv
63
64 """
65 print读取的tfrecords文件，这个是逐行读取，其中由于tensorflow不支持转换为string，采用了bytes.decode转换
66 """
67 def printRecords(user_id, age, sex, user_lv_cd, user_reg_dt):
68     init = tf.global_variables_initializer()
69     sess = tf.Session()
70     sess.run(init)
71     threads = tf.train.start_queue_runners(sess=sess)
72     for i in range(10):
73         val_user_id, val_age, val_sex, val_user_lv_cd, val_user_reg_dt = sess.run(
74             [user_id, age, sex, user_lv_cd, user_reg_dt])

```



返回顶部



喜欢



收藏



评论



分享

```

75     print(val_user_id, bytes.decode(val_age), val_sex, val_user_lv_cd, bytes.decode(val_user_reg_dt))
76
77 """
78 print读取的tfrecords文件，这个批量读取(分为批量打乱读取和批量读取)，一般实际训练模型采用这种读取方式。
79 """
80 def calcRecords(m_user_id, m_age, m_sex, m_user_lv_cd, m_user_reg_dt):
81     user_id = tf.cast(m_user_id,tf.int64)
82     age = tf.cast(m_age,tf.string)
83     sex = tf.cast(m_sex,tf.int64)
84     user_lv_cd = tf.cast(m_user_lv_cd,tf.int64)
85     user_reg_dt = tf.cast(m_user_reg_dt,tf.string)
86
87     init = tf.global_variables_initializer()
88     sess = tf.Session()
89     sess.run(init)
90     # tf.train.shuffle_batch批量打乱并数据
91     val_user_id, val_age, val_sex, val_user_lv_cd, val_user_reg_dt = tf.train.shuffle_batch([user_id, age, sex, user_lv_cd, user_reg_dt],
92         batch_size=10,
93         capacity=2000,
94         min_after_dequeue=1000,
95         num_threads=12)
96
97     # tf.train.batch批量取数据
98     # val_user_id, val_age, val_sex, val_user_lv_cd, val_user_reg_dt = tf.train.batch([user_id, age, sex, user_lv_cd, user_reg_dt],
99         # batch_size=10,
100        # capacity=2000,
101        # num_threads=12)
102     threads = tf.train.start_queue_runners(sess=sess)
103     for i in range(10):
104         p_user_id, p_age, p_sex, p_user_lv_cd, p_user_reg_dt = sess.run([val_user_id, val_age, val_sex, val_user_lv_cd, val_user_reg_dt])
105         print(p_user_id, p_sex, p_user_lv_cd)
106
107
108 def main(unused_argv):
109     """
110     csv文件,格式：
111     user_id,age,sex,user_lv_cd,user_reg_dt

```



返回顶部



喜欢



收藏



评论



分享

```
112 1,46-55岁,0,5,2004-10-12
113 2,19-25岁,2,3,2013-04-10
114 3,26-35岁,2,4,2016-01-26
115 4,-1,2,1,2016-01-26
116 5,-1,2,3,2016-01-26
117 6,-1,2,1,2016-01-26
118 7,19-25岁,2,3,2016-01-26
119 8,26-35岁,2,3,2016-01-26
120 9,26-35岁,0,4,2013-04-10
121 10,26-35岁,0,3,2016-01-26
122 """"
123 # save train TfRecords文件
124 train_data_set = getDataSet(FLAGS.train_path)
125 saveTfRecords(train_data_set, 'train')
126
127 # save train TfRecords文件
128 test_data_set = getDataSet(FLAGS.test_path)
129 saveTfRecords(test_data_set, 'test')
130
131 # read train TfRecords文件
132 user_id, age, sex, user_lv_cd, user_reg_dt = readTfRecords("train")
133 calcRecords(user_id, age, sex, user_lv_cd, user_reg_dt)
134
135 # read test TfRecords文件
136 user_id, age, sex, user_lv_cd, user_reg_dt = readTfRecords("test")
137 printRecords(user_id, age, sex, user_lv_cd, user_reg_dt)
138
139 if __name__ == '__main__':
140     parser = argparse.ArgumentParser()
141     parser.add_argument(
142         '--train_path',
143         type=str,
144         default=r"E:\jdata_user\JData_User.csv",
145         help='read train file'
146     )
147     parser.add_argument(
148         '--test_path',
```



返回顶部

```

149     type=str,
150     default=r"E:\jdata_user\JData_User_Test.csv",
151     help='read test file'
152 )
153 parser.add_argument(
154     '--dir_path',
155     type=str,
156     default=r"E:\jdata_user",
157     help='Directory to save data files and write the converted result'
158 )
159 FLAGS, unparsed = parser.parse_known_args()
160 tf.app.run(main=main, argv=[sys.argv[0]] + unparsed)

```



喜欢



收藏



注：



(1) 上述代码片段是对演示TFRecords文件的生成和读取显示过程，其中包含了对字符串的特殊处理过程，由于评论 tf.train.Feature 不支持string类型，所以save时候把字符串转换为byte后在读取时候再转换为string显示。



(2) 在实际tensorflow使用场景中，一般字符串不参加运算，所以在生成TFRecords文件不建议包含字符串变量（如果必须包含字符串建议转化为词向量参与运算）

分享

(3) 本代码片段所涉及文件结果如下图

名称	修改日期	类型	大小
JData_User.csv	2017-05-27 18:04	Microsoft Excel ...	6,831 KB
JData_User_Test.csv	2017-06-01 13:07	Microsoft Excel ...	308 KB
test.tfrecords	2017-06-01 13:14	TFRECORDS 文件	1,291 KB
train.tfrecords	2017-06-01 13:14	TFRECORDS 文件	28,970 KB



返回顶部

(4) 一般读取数据并训练模型采用上述代码片段中的calcRecords函数的方式(打乱数据顺序并批量读取), 读取显示后的效果如下:

```
[506 565 674 864 990 75 628 83 874 931] [2 2 2 2 2 2 2 2 0 1] [3 3 2 2 4 2 2 4 4 4]
[ 899 1064 991 475 253 979 276 791 513 898] [2 2 2 0 2 0 2 2 2 0] [5 2 2 2 4 5 1 3 4 4]
[1098 825 535 336 1117 1000 25 290 390 729] [2 2 2 2 2 2 2 2 0 2] [2 2 2 2 1 3 3 2 2 4]
[ 174 992 5 281 986 474 1126 371 52 16] [2 2 2 2 2 0 2 2 2 2] [3 1 3 1 4 5 2 2 3 2]
[884 504 396 939 376 134 831 922 621 629] [2 2 2 2 2 2 2 0 2 1] [3 1 2 2 2 3 2 5 2 3]
[1086 588 3 1055 201 289 485 692 445 551] [0 0 2 2 2 0 0 2 2 2] [5 2 4 2 3 5 4 2 2 2]
[ 193 927 877 1002 892 572 845 1047 1263 943] [2 2 0 2 2 2 2 2 2 1] [2 3 3 1 2 2 2 2 3 2]
[1197 554 448 465 503 172 762 204 764 622] [2 0 2 2 2 2 2 0 2 2] [2 4 3 2 2 2 2 2 4 3]
[691 896 65 58 328 97 683 48 938 484] [0 2 0 2 2 0 2 2 2 1] [3 1 4 2 3 4 3 1 4 1]
[100 901 718 893 921 316 115 263 159 984] [2 0 2 0 1 2 0 2 2 2] [2 3 4 3 4 2 2 1 2 2]
```

Process finished with exit code 0

<http://blog.csdn.net/oTengYue>

- 喜欢
- 收藏
- 评论
- 分享

扩展: tensorflow中的数据类型

除了维度, tensor 有一个数据类型属性。你可以为一个张量指定下列数据类型中的任何一个类型:

数据类型	python类型	描述
DT_FLOAT	tf.float32	32位浮点数
DT_DOUBLE	tf.float64	64位浮点数
DT_INT64	tf.int64	64位有符号整型
DT_INT32	tf.int32	32位有符号整型
DT_INT16	tf.int16	16位有符号整型
DT_INT8	tf.int8	8位有符号整型
DT_UINT8	tf.uint8	8位无符号整型
DT_STRING	tf.string	可变长度的字节数组, 每一个张量元素都是一个字节数组
DT_BOOL	tf.bool	布尔型
DT_COMPLEX64	tf.complex64	由32位浮点数组成的复数: 实数和虚数
DT_QINT32	tf.qint32	用于量化Ops的32位有符号整型
DT_QINT8	tf.qint8	用于量化Ops的8位有符号整型
DT_QUINT8	tf.quint8	用于量化Ops的8位无符号整型

<http://blog.csdn.net/oTengYue>

版权声明：本文为博主原创文章，未经博主允许不得转载。

 举报

标签：tensorflow (<http://so.csdn.net/so/search/s.do?q=tensorflow&t=blog>) /

TFRecords (<http://so.csdn.net/so/search/s.do?q=TFRecords&t=blog>) /



喜欢



收藏



评论



分享

### 相关文章推荐

#### tensorflow中tfrecords文件的save和read (/otengyue/article/details/72830262)

在tensorflow程序中，推荐使用tensorflow内定标准格式——TFRecords，本文是对csv生成TFRecords文件并读取显示的代码实现



oTengYue (<http://blog.csdn.net/oTengYue>) 2017-06-01 13:14  328

#### Notes on tensorflow (八) read tfrecords with slim (/weixin\_35653315/article/details/71023596)

[http://blog.csdn.net/weixin\\_35653315/article/details/71015845](http://blog.csdn.net/weixin_35653315/article/details/71015845) 中描述了将pascal voc数据集转换tfrecord的过程。本文则通过S...




返回顶部



 weixin\_35653315 ([http://blog.csdn.net/weixin\\_35653315](http://blog.csdn.net/weixin_35653315)) 2017-04-30 16:21  1374




## 精选：深入理解 Docker 内部原理及网络配置 ([http://edu.csdn.net/huiyiCourse/detail/563?utm\\_source=blog10](http://edu.csdn.net/huiyiCourse/detail/563?utm_source=blog10))

 多令人称道的功能，如 Docker 的镜像管理。然而，Docker的网络一直以来都比较薄弱，所以我们有必要深入了解Docker的喜欢网络知识，以满足更高的网络需求。



## 收藏 **Tensorflow读取数据2-tfrecord (/u010911921/article/details/70991194)**

 用自己的数据集创建Tensorflow的标准格式TFRecords

评论



 u010911921 (<http://blog.csdn.net/u010911921>) 2017-04-29 22:47  359



分享

## **Tensorflow高阶读写教程 (/u010223750/article/details/70482498)**

前言tensorflow提供了多种读写方式，我们最常见的就是使用tf.placeholder()这种方法，使用这个方法需要我们提前处理好数据格式，不过这种处理方法也有缺陷：不便于存储和不利于分布式处理...

 u010223750 (<http://blog.csdn.net/u010223750>) 2017-04-24 21:54  1852

## **7.2 TensorFlow笔记(基础篇): 生成TFRecords文件 (/fontthrone/article/details/76727412)**

前言在TensorFlow中进行模型训练时,在官网给出的三种读取方式,中最好的文件读取方式就是将利用队列进行文件读取,而且步骤有两步: 1. 把样本数据写入TFRecords二进制文件 2. ...

 FontThrone (<http://blog.csdn.net/FontThrone>) 2017-08-05 17:18  242

  
返回顶部

## Tensorflow中使用TFRecords高效读取数据--结合NLP数据实践 (/liuchonge/article/details/73649251)

之前一篇博客在进行论文仿真的时候用到了TFRecords进行数据的读取操作，但是因为当时比较忙，所以没有进行深入学习。这两天看了一下，决定写篇博客专门结合该代码记录一下TFRecords的相关操作。 ...



liuchonge (<http://blog.csdn.net/liuchonge>) 2017-06-23 20:15 7397



喜欢

## Tensorflow Save (/efforever/article/details/75137032)



保存为四个文件： my-model.ckpt.meta my-model.ckpt.data-\* my-model.ckpt.index checkpoint 图结构.meta与变量值....

收藏



efforever (<http://blog.csdn.net/efforever>) 2017-07-14 21:52 62



评论

## Tensorflow 中Tfrecords的使用心得 (/u014802590/article/details/68495238)



分享

这篇博客主要讲了如何用Tensorflow中的标准数据读取方式的简单的实现对自己数据的读取操作。



u014802590 (<http://blog.csdn.net/u014802590>) 2017-03-30 20:10 517

## tensorflow中cifar-10文档的Read操作 (/u010223750/article/details/53244744)



前言 在tensorflow的官方文档中得卷积神经网络一章，有一个使用cifar-10图片数据集的实验，搭建卷积神经网络倒不难，但是那个cifar10\_input文件着实让我费了一番心思。配合着官方...



u010223750 (<http://blog.csdn.net/u010223750>) 2016-11-20 22:29 1586

## TFRecords 文件的生成和读取 (/u012222949/article/details/72875281)

TensorFlow提供了TFRecords的格式来统一存储数据，理论上，TFRecords可以存储任何形式的数据。 TFRecords文件中的数据都是通过tf.train.Example ...



 u012222949 (<http://blog.csdn.net/u012222949>) 2017-06-06 10:09  1368

### tensorflow模型参数保存和加载问题 (/u014659656/article/details/53954793)

tensorflow模型参数保存和加载问题终于找到bug原因！记一下；还是不熟悉平台的原因造成的！Q：为什么会出现两个模型对象在同一个文件中一起运行，当直接读取他们分开运行时训练出来的模型会出错，...



喜欢

 u014659656 (<http://blog.csdn.net/u014659656>) 2016-12-31 12:33  4036



收藏

### canvas的save和restore (/duantihi/article/details/50466900)

经过画图琢磨，终于理解 canvas的save和restore的用意了，上效果图和：代码 左下角画一个圆、中间画一个向右箭头、右上角画一个圆 ps: 使用了restore的效果 ps: 不使...



评论



分享

 duantihi (<http://blog.csdn.net/duantihi>) 2016-01-06 09:55  1289

### 学习TensorFlow，保存学习到的网络结构参数并调用 (/helei001/article/details/51697658)

在深度学习中，不管使用那种学习框架，我们会遇到一个很重要的问题，那就是在训练完之后，如何存储学习到的深度网络的参数？在测试时，如何调用这些网络参数？针对这两个问题，本篇博文主要探索TensorFlow...

 helei001 (<http://blog.csdn.net/helei001>) 2016-06-17 09:55  10974

### 7.1 TensorFlow笔记(基础篇):加载数据之预加载数据与填充数据 (/fontthroner/article/details/76727466)

TensorFlow加载数据TensorFlow官方共给出三种加载数据的方式: 1. 预加载数据 2. 填充数据 预加载数据的缺点: 将数据直接嵌在数据流图中,当训练数据较大时,很消耗内存....



返回顶部



FontThrone (<http://blog.csdn.net/FontThrone>) 2017-08-05 17:24 198

## C语言循环输入字符串#结束写入文件 (/android\_hasen/article/details/9735329)

```
#include #include #include #include #include #define BUF_SIZE 128 void my_read(int fd) { int s...
```



u011430511 (<http://blog.csdn.net/u011430511>) 2013-08-03 21:00 1101

喜欢



Hibernate深入理解----03Session缓存核心方法 ( save()和persist() , get()和load(),update()和saveOrUpdate,delete()和evil() ) (/ochangwen/article/details/52583897)

收藏



参考代码下载github:<https://github.com/changwensir/java-ee/tree/master/hibernate4> 一、持久化对象的状态 1).临时对象 ( Tr  
ans...

评论



oChangWen (<http://blog.csdn.net/oChangWen>) 2016-09-19 11:37 1015

分享

## NodeIO\_readFile&readFileSync (/qq\_27932435/article/details/75570574)

Node中部分文件方法对比readFile和readFileSync的比较 : var fs = require("fs");var data = fs.readFileSync('main....



qq\_27932435 ([http://blog.csdn.net/qq\\_27932435](http://blog.csdn.net/qq_27932435)) 2017-07-20 17:40 69

## 实现read文件 (/feike24/article/details/52443386)

```
read #include ssize_t read ( int fd, // 文件描述符 void* buf, // 缓冲区 size_t count // 期...
```



feike24 (<http://blog.csdn.net/feike24>) 2016-09-05 20:32 68

## Canvas的save和restore (/xajax/article/details/6862085)

返回顶部

在创建新的控件或修改现有的控件时，我们都会涉及到重写控件或View的onDraw方法。 onDraw方法会传入一个Canvas对象，它是你用来绘制控件视觉界面的画布。 ...


 xajax (<http://blog.csdn.net/xajax>) 2011-10-11 11:34  943

### Canvas的save和restore (/mowen1111/article/details/8960085)



在创建新的控件或修改现有的控件时，我们都会涉及到重写控件或View的onDraw方法。 onDraw方法会传入一个Canvas对象，它是你用来绘制控件视觉界面的画布。 ...



mowen1111 (<http://blog.csdn.net/mowen1111>) 2013-05-22 14:08  432

收藏



评论



分享



返回顶部