

[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>)

目录



喜欢



收藏



评论



分享

# tensorflow学习笔记（四十）：tensorflow语音识别 及 python音频处

## 理库

原创 2017年05月04日 20:22:43

5038

1

8

## tensorflow 语音识别

最近在做语音识别的项目，现在项目告一段落，就把最近碰到的东西做一个总结。

## python中关于语音处理的库

- `scipy.io.wavfile`
- `python_speech_features`
- 读取wav文件

```
1 import scipy.io.wavfile as wav
2 fs, audio = wav.read(file_name)
```

- 对读取的音频信息求MFCC（Mel频率倒谱系数）

ke1th (<http://blog.csdn.net/>...[+ 关注](#)<http://blog.csdn.net/u012436149>

码云

原创

128

粉丝

292

喜欢

0

未开通

[https://github.com/ke1th/utm\\_source](https://github.com/ke1th/utm_source)

### 他的最新文章

[更多文章 \(http://blog.csdn.net/u012436149\)](http://blog.csdn.net/u012436149)

softmax交叉熵与最大似然估计

[\(/u012436149/article/details/78006552\)](http://blog.csdn.net/u012436149/article/details/78006552)tensorflow学习笔记（四十三）：使用  
tfdbg 来 debug[\(/u012436149/article/details/77479425\)](http://blog.csdn.net/u012436149/article/details/77479425)

一段代码搞懂 gpu memory

[\(/u012436149/article/details/77221417\)](http://blog.csdn.net/u012436149/article/details/77221417)

pytorch: 如何优雅的将 int list 转成 one-hot形式

[\(/u012436149/article/details/77017832\)](http://blog.csdn.net/u012436149/article/details/77017832)

```
1 from python_speech_features import mfcc
2 from python_speech_features import delta
3 #求MFCC
4 processed_audio = mfcc(audio, samplerate=fs)
5 #求差分(一阶,二阶)
6 delta1 = delta(processed_audio, 1)
7 delta2 = delta(processed_audio, 2)
```



## 目录 pydub



github 项目地址 (<https://github.com/jiaaro/pydub>)



有了这个库,做音频的数据增强就容易多了.关于使用方法可以阅读 github 上的文档,这里只对 raw\_data 做一些说明.



收藏

```
1 raw_audio_data = sound.raw_data
```



评论

raw\_audio\_data 中包含的是 音频数据的 bytestring ,但是如果我们想对音频数据做 MFCC ,那么我们应该怎么办呢?



分享

```
1 audio = np.fromstring(raw_audio_data, dtype=np.int16)
2 #此时audio是一个一维的ndarray,如果音频是双声道,
3 #我们只需要对其进行reshape就可以了
4 audio = np.reshape(audio, [-1, 2])
5
6 # 然后就可以使用python_speech_features做进一步操作了
```

## tensorflow中做语音识别会碰到的API

这个部分包括了 SparseTensor , sparse\_tensor\_to\_dense , edit\_distance 。

### SparseTensor(indices, values, dense\_shape)

- indices: 一个2D的 int64 Tensor ,shape为 (N, ndims) ,指定了 sparse tensor 中的索引, 例如: indices=[[1,3], [2,4]] 说明, dense tensor 中对索引为 [1,3], [2,4] 位置的元素的值不为0.

编辑推荐

最新专栏

Tensorflow一些常用基本概念与函数 (1...

Tensorflow中使用 LSTM + CTC 的例子...

tensorflow学习笔记(十七):name&variab...

### 在线课程



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



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



返回顶部

- values: 一个 1D tensor , shape 为 (N) 用来指定索引处的值. For example, given indices=[[1,3], [2,4]], the parameter values=[18, 3.6] specifies that element [1,3] of the sparse tensor has a value of 18, and element [2,4] of the tensor has a value of 3.6.
- dense\_shape: 一个1D的 int64 tensor ,形状为 ndims ,指定 dense tensor 的形状.

相对应的有一个 tf.sparse\_placeholder ,如果给这个 sparse\_placeholder 喂数据呢?

```
1 sp = tf.sparse_placeholder(tf.int32)
2
3 with tf.Session() as sess:
4     #就这么喂就可以了
5     feed_dict = {sp:(indices, values, dense_shape)}
```

tensorflow 中目前没有API提供denseTensor->SparseTensor转换

`tf.sparse_tensor_to_dense(sp_input, default_value=0, validate_indices=True, name=None)`

把一个 SparseTensor 转化为 DenseTensor .

- sp\_input: 一个 SparseTensor .
- default\_value:没有指定索引的对应的默认值.默认为0.
- validate\_indices: 布尔值.如果为 True 的话,将会检查 sp\_input 的 indices 的 lexicographic order 和是否有重复.
- name: 返回tensor的名字前缀.可选.

`tf.edit_distance(hypothesis, truth, normalize=True, name='edit_distance')`



目录



喜欢



收藏



评论



分享



返回顶部

计算序列之间的 Levenshtein 距离

- hypothesis: SparseTensor ,包含序列的假设.
- truth: SparseTensor , 包含真实序列.
- normalize: 布尔值,如果值 True 的话,求出来的 Levenshtein 距离除以真实序列的长度. 默认为 True
- name: operation 的名字,可选.



目录

返回值:



返回值是一个 R-1 维的 DenseTensor .包含着每个 Sequence 的 Levenshtein 距离.

喜欢

SparseTensor 所对应的 DenseTensor 是一个多维的 Tensor ,最后一维看作序列.



收藏

## CTCloss



评论

现在用深度学习做语音识别,基本都会在最后一层用 CTCloss ,这个 loss 自己实现起来还是有点费劲,不过,幸运的是, tensorflow 中已经有现成的 API 了,我们只需调用即可。



分享

```
tf.nn.ctc_loss(labels, inputs, sequence_length,
preprocess_collapse_repeated=False,
ctc_merge_repeated=True)
```

此函数用来计算 ctc loss .

- labels:是一个 int32 的 SparseTensor , labels.indices[i, :] == [b, t] 表示 labels.values[i] 保存着 (batch b, time t) 的 id .
- inputs:一个 3D Tensor (max\_time \* batch\_size \* num\_classes) .保存着 logits .(通常是 RNN 接上一个线性神经元的输出)



返回顶部

- sequence\_length: 1-D int32 向量, size 为 [batch\_size] . 序列的长度.此 sequence\_length 和用在 dynamic\_rnn 中的sequence\_length是一致的,使用来表示 rnn 的哪些输出不是 pad 的.
- preprocess\_collapse\_repeated:设置为 True 的话, tensorflow 会对输入的 labels 进行预处理,连续重复的会被合成一个.
- ctc\_merge\_repeated: 连续重复的是否被合成一个



返回值:

目录 一个 1-D float Tensor, size 为 [batch] , 包含着负的  $\log p$ .加起来即为 batch loss .



喜欢

`tf.nn.ctc_greedy_decoder(inputs, sequence_length, merge_repeated=True)`



收藏

上面的函数是用在训练过程中,专注与计算 loss ,此函数是用于 inference 过程中,用于解码.



评论

- inputs:一个 3D Tensor (max\_time \* batch\_size \* num\_classes) .保存着 logits .(通常是 RNN 接上一个线性神经元的输出)



分享

- sequence\_length: 1-D int32 向量, size 为 [batch\_size] . 序列的长度.此 sequence\_length 和用在 dynamic\_rnn 中的sequence\_length是一致的,使用来表示 rnn 的哪些输出不是 pad 的.

返回值:

一个 tuple (decoded, log\_probabilities)

- decoded: 一个只有一个元素的哦 list . decoded[0] 是一个 SparseTensor ,保存着解码的结果.
  - decoded[0].indices: 索引矩阵,size为 (total\_decoded\_outputs \* 2) ,每行中保存着 [batch, time] .
  - decoded[0].values: 值向量, size 为 (total\_decoded\_outputs) .向量中保存的是解码的类别.
  - decoded[0].shape: 稠密 Tensor 的 shape , size为 (2) . shape 的值为 [batch\_size, max\_decoded\_length] .
- log\_probability: 一个浮点型矩阵 (batch\_size\*1) 包含着序列的log 概率.

```
tf.nn.ctc_beam_search_decoder(inputs, sequence_length,  
beam_width=100, top_paths=1, merge_repeated=True)
```

另一种寻路策略。和上面那个差不多。

知道这些，就可以使用 tensorflow 搭建一个简单的语音识别应用了。

## 参考资料

目录



喜欢

[https://www.tensorflow.org/api\\_docs/python/tf/nn/ctc\\_loss](https://www.tensorflow.org/api_docs/python/tf/nn/ctc_loss) ([https://www.tensorflow.org/api\\_docs/python/tf/nn/ctc\\_loss](https://www.tensorflow.org/api_docs/python/tf/nn/ctc_loss))

[https://www.tensorflow.org/api\\_docs/python/tf/nn/ctc\\_greedy\\_decoder](https://www.tensorflow.org/api_docs/python/tf/nn/ctc_greedy_decoder)

([https://www.tensorflow.org/api\\_docs/python/tf/nn/ctc\\_greedy\\_decoder](https://www.tensorflow.org/api_docs/python/tf/nn/ctc_greedy_decoder))



收藏

[https://www.tensorflow.org/api\\_docs/python/tf/nn/ctc\\_beam\\_search\\_decoder](https://www.tensorflow.org/api_docs/python/tf/nn/ctc_beam_search_decoder)

([https://www.tensorflow.org/api\\_docs/python/tf/nn/ctc\\_beam\\_search\\_decoder](https://www.tensorflow.org/api_docs/python/tf/nn/ctc_beam_search_decoder))



评论

<http://stackoverflow.com/questions/38059247/using-tensorflows-connectionist-temporal-classification-ctc-implementation>

(<http://stackoverflow.com/questions/38059247/using-tensorflows-connectionist-temporal-classification-ctc-implementation>)

[https://www.tensorflow.org/versions/r0.10/api\\_docs/python/nn/connectionist\\_temporal\\_classification\\_\\_ctc\\_](https://www.tensorflow.org/versions/r0.10/api_docs/python/nn/connectionist_temporal_classification__ctc_)

([https://www.tensorflow.org/versions/r0.10/api\\_docs/python/nn/connectionist\\_temporal\\_classification\\_\\_ctc\\_](https://www.tensorflow.org/versions/r0.10/api_docs/python/nn/connectionist_temporal_classification__ctc_))



分享

版权声明：本文为博主原创文章，转载请注明出处。

举报

本文已收录于以下专栏：tensorflow学习笔记 (<http://blog.csdn.net/column/details/13300.html>)



返回顶部



(H\_o\_w\_e) 2017-08-21 20:13

1楼

目录 有没源码？我参照这里（[http://blog.csdn.net/sinat\\_30665603/article/details/74897891#reply](http://blog.csdn.net/sinat_30665603/article/details/74897891#reply)）编译的，结果一句话只能识别出一个字，但不正确：



回复

喜欢



收藏

### 相关文章推荐



评论 **tensorflow学习笔记（一）：命令行参数 (/u012436149/article/details/52870069)**



全局环境下编写代码import tensorflow as tf flags = tf.flags logging = tf.logging flags.DEFINE\_string("para\_nam...



分享 u012436149 (<http://blog.csdn.net/u012436149>) 2016-10-20 10:54 8060

### Tensorflow一些常用基本概念与函数（1） (/lenbow/article/details/52152766)

摘要：本文主要对tf的一些常用概念与方法进行描述。



lenbow (<http://blog.csdn.net/lenbow>) 2016-08-08 16:00 47536

**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))**

返回顶部



网络绝对是任何系统的核心，对于容器而言也是如此。Docker 作为目前最火的轻量级容器技术，有很多令人称道的功能，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的网络知识，以满足更高的网络需求。

## Tensorflow中使用 LSTM + CTC 的例子 (/zchang81/article/details/77043628)



目录



喜欢



收藏



评论



分享

zchang81 (<http://blog.csdn.net/zchang81>) 2017-08-10 11:20 322

## tensorflow学习笔记(十七):name&variable scope (/u012436149/article/details/53081454)

name&variable scope水平有限,如有错误,请指正!在tensorflow中,有两个scope, 一个是name\_scope一个是variable\_scope,这两个scope到底有什么...

u012436149 (<http://blog.csdn.net/u012436149>) 2016-11-08 13:49 11352

## tensorflow学习笔记(十五): variable scope (/u012436149/article/details/53018924)

variable scopetensorflow 为了更好的管理变量,提供了variable scope机制 官方解释: Variable scope object to carry default...

u012436149 (<http://blog.csdn.net/u012436149>) 2016-11-03 10:25 7135

## ( ZT ) 贝索斯:云计算是亚马逊的一粒种子 (/decipheringcloud/article/details/6461700)

4月27日，亚马逊发布了让投资者略为失望的第一财季财报。报告显示，亚马逊第一季度净销售额为98.6亿美元，比去年同期的71.3亿美元增长 38%；但净利润为2.01亿美元，比去年同期的2.99亿美元下...


DecipheringCloud (<http://blog.csdn.net/DecipheringCloud>) 2011-06-02 16:23 179

返回顶部



## 如何用Tensorflow开发一个简单的语音识别器 (/lc0817/article/details/53579366)

在这个视频中，我们将使用Tensorflow机器学习库，用20行Python代码创建一个超简单的语音识别器。一起回顾下语音识别研究的历史，然后解释如何使用深度学习建立自己的语音识别系统。视频连接...

 lc0817 (<http://blog.csdn.net/lc0817>) 2016-12-12 09:23 5530



## CSDN上看到的一些小问题 (/mamingtyut/article/details/6741207)

目录



//===== //1:2011-09-01 int tmp = printf(...



mamingtyut (<http://blog.csdn.net/mamingtyut>) 2011-09-01 22:56 30



## 基于Tensorflow的VCTK语音识别例子测试 (/sparkexpert/article/details/71170709)



语音识别是深度学习首先攻克的几个领域之一。传统的基于HMM等的语音识别精度一直比较受限。但是深度学习还是给语音识别的精度带来了一个飞跃性的提高。本文在网上找了段代码实现了下，感觉非常简单就可以复现。 ...

评论




sparkexpert (<http://blog.csdn.net/sparkexpert>) 2017-05-04 15:12 881

分享

## TensorFlow 中文语音识别 (/sinat\_30665603/article/details/74897891)

本文转载自 <http://blog.topspeedsnail.com/archives/10696> 数据集下载参见该文。其中下面的代码进行了一些小小的调整。其中包含缩进、版本方面(作...

 sinat\_30665603 ([http://blog.csdn.net/sinat\\_30665603](http://blog.csdn.net/sinat_30665603)) 2017-07-09 18:09 745

## Kaldi语音识别学习笔记（一） (/xiaoseyihe/article/details/73918220)

新开通的CSDN博客。对工作和学习中的一些东东和大家一起分享，欢迎指点。闲话不多说。之前搞了一段时间htk语音识别，现在流行kaldi，自己也开始学习一下，在分享的过程中，如果我的理...



返回顶部



xiaoseyihe (<http://blog.csdn.net/xiaoseyihe>) 2017-06-29 16:33 61

## 【神经网络学习笔记】语音识别-matlab语音信号预处理 (/zjccoder/article/details/37873499)

要让电脑识别语音信号，必须经过一系列的处理。



zjccoder (<http://blog.csdn.net/zjccoder>) 2014-07-16 12:43 2772

目录



## tensorflow学习笔记（四）：TensorBoard (/woaidapaopao/article/details/72911763)

喜欢一、TensorBoard 我们上一节介绍过简单的TensorFlow手写字符识别程序了，但是在实际中调参和观察flow的状态是很有必要的，于是TensorFlow便有了TensorBoard帮...



收藏 woaidapaopao (<http://blog.csdn.net/woaidapaopao>) 2017-06-10 11:24 254



## 评论 TensorFlow学习笔记（二） (/tan\_chi\_she/article/details/70140958)



之前安装TensorFlow之后，先简单测试了一下，了解了TensorFlow的基本模型以及session接口，接下来需要知道这个TensorFlow到底是怎么执行运行的，它的整个流程是什么，它的框架...



tan\_chi\_she ([http://blog.csdn.net/tan\\_chi\\_she](http://blog.csdn.net/tan_chi_she)) 2017-04-12 11:22 136

## Sphinx语音识别学习记录（四）-小范围语音中文识别 (/yuanya/article/details/9229499)

CMUSphinx系列目录 关于中文识别 网上很多都是学习自下面的博客，Sphinx武林秘籍(上) 但很多人照做之后纷纷表示识别不出来，我们猜测是某个参数设定错误或者版本变化导致的 ht...



yuanya (<http://blog.csdn.net/yuanya>) 2013-07-02 23:42 1073

## tensorflow学习笔记 (/wqqgo/article/details/75675323)



返回顶部

tensorflow学习笔记按照《TensorFlow：实战Google深度学习框架》一书学习的tensorflow，书中使用的是0.9.0版本，而我安装的是1.2.1，出现了一些问题：1、使用pi...



wqqGo (<http://blog.csdn.net/wqqGo>) 2017-07-21 21:59 162



## TensorFlow学习笔记 (/y276455978/article/details/76651323)



笔记课程链接：<http://edu.csdn.net/huiyiCourse/attend/433> 一、TensorFlow 安装pip install tensorflow二、开发环境起内置的



W...



y276455978 (<http://blog.csdn.net/y276455978>) 2017-08-03 21:42 25



## 语音识别原理 (/smartempire/article/details/24411607)



本文 简要给大家介绍一下语音怎么变文字的。首先说一下作为输入的时域波形。我们知道声音实际上是一种波。常见的mp3、wmv等格式都是压缩格式，必须转成非压缩的纯波形文件，比如Windows ...



评论



u010944555 (<http://blog.csdn.net/u010944555>) 2014-04-24 14:29 3286

分享

## 音频录制与iOS10语音识别 (/bluecat\_1128/article/details/52877270)

iOS10新特性 语音识别 音频录制



blueCat\_1128 ([http://blog.csdn.net/blueCat\\_1128](http://blog.csdn.net/blueCat_1128)) 2016-10-20 21:36 1122