

[深度學習] TensorFlow中模型的freeze_graph



一隻淡水龜

新加坡國立大學計算機碩士

6人讚同了該文章

在TensorFlow中模型的保存和調用，相信大家都不會陌生，使用關鍵語句`saver = tf.train.Saver()`和`saver.save()`就可以完成。

但是，不知道大家是否了解，tensorflow通過checkpoint這一種格式文件，是將模型的結構和權重數據分開保存的，這就造成了一些使用場景下的不方便。

所以，我們需要一種方式將模型結構和權重數據合併在一個文件中，tensorflow提供了`freeze_graph`函數和pb文件格式，來解決這一問題。

這些模型文件是做什麼的

```
checkpoint
events.out.tfevents.1547198372.pci-SYS-7048GR-TR
model-20190111-171825.ckpt-322000.data-00000-of-00001
model-20190111-171825.ckpt-322000.index
model-20190111-171825.ckpt-322000.meta
```

在save之後，模型會保存在ckpt文件中，checkpoint文件保存了一個目錄下所有的模型文件列表，events文件是給可視化工具tensorboard用的。

和保存的模型直接相關的是以下這三個文件：

- .data文件保存了當前參數值
- .index文件保存了當前參數名
- .meta文件保存了當前圖結構

當你使用`saver.restore()`載入模型時，你用的就是這一組的三個checkpoint文件。

但是，當我們需要將模型和權重整合成一個文件時，我們就需要以下的操作了。

如何使用freeze_graph生成PB文件

tensorflow提供了`freeze_graph`這個函數來生成pb文件。以下的代碼塊可以完成將checkpoint文件轉換成pb文件的操作：

載入你的模型結構，
提供checkpoint文件地址
使用`tf.train.writegraph`保存圖，這個圖會提供給`freeze_graph`使用
使用`freeze_graph`生成pb文件

```
import tensorflow as tf
from tensorflow.python.tools import freeze_graph

# network是你自己定义的模型
import network
```

模型的checkpoint文件地址

```
ckpt_path = "./ckpt_model/model-20190403-164504.ckpt-2050"
```

```
def main():
```

```
    tf.reset_default_graph()
```

```
    x = tf.placeholder(tf.float32, shape=[None, 224,
```

flow是模型的输出

```
    flow = network(x)
```

设置输出类型以及输出的接口名字, 为了之后的调用pb的时

```
    flow = tf.cast(flow, tf.int8, 'out')
```

```
with tf.Session() as sess:
```

保存图, 在./pb_model文件夹中生成model.pb文

model.pb文件将作为input_graph给到接下来的

```
    tf.train.write_graph(sess.graph_def, './p'
```

把图和参数结构一起

```
    freeze_graph.freeze_graph(  
        input_graph='./pb_model/model.pb'
```

```
        input_saver='',
```

```
        input_binary=False,
```

```
        input_checkpoint=ckpt_path,
```

```
        output_node_names='out',
```

```
        restore_op_name='save/restore_all'
```

```
        filename_tensor_name='save/Const':
```

```
        output_graph='./pb_model/frozen_r'
```

```
        clear_devices=False,
```

```
        initializer_nodes='')
```

```
)
```

```
print("done")
```

```
if __name__ == '__main__':  
    main()
```

在以上的程序運行之後，./pb_model/文件夾中就會出現frozen_model.pb文件，這是我們可以使用的模型結構和權重整合過的pb文件。

freeze_graph總共有11個參數，以下逐一介紹下，供大家參考：

input_graph: 模型文件，可以是二進制的pb文件，或文本的meta文件，用input_binary來指定區分。我們的例子中，使用了二進制的pb文件，對應input_binary就是False

input_saver: Saver解析器，主要用於版本不兼容時使用。通常為空，為空時用當前版本的Saver

input_binary: 配合input_graph用，為true時，input_graph為二進制，為false時，input_graph為文件。默認值是False

input_checkpoint: checkpoint文件地址

output_node_names: 輸出節點的名字，有多個時用逗號分開，我們的輸出節點是'out'，這是我們使用flow = tf.cast(flow, tf.int8, 'out')將模型的輸出節點命名為out。如果沒有這一步的操作，我們可以找到模型的輸出節點名是什麼，並且在這一參數中對應。

restore_op_name: 從模型恢復節點的名字，一般使用默認：
save/restore_all

filename_tensor_name: 一般使用默認: save/Const:0

output_graph: 用來保存整合後的模型輸出文件, 即pb文件的保存地址

clear_devices: 指定是否清除訓練時節點指定的運算設備 (如cpu、gpu、tpu。cpu是默認), 默認True

initializer_nodes: 默認空。權限加載後, 可通過此參數來指定需要初始化的節點, 用逗號分隔多個節點名字。

variable_names_blacklist: 默認空。變量黑名單, 用於指定不用恢復值的變量, 用逗號分隔多個變量名字。