



【译】在Windows下安装XGBoost

📅 2017-02-21 | 📁 教程 | 👁 6814

XGBoost是Gradient Boosting算法的一种高级实现，在Kaggle competitions上崭露头角。下面就对XGBoost在Windows上的安装作一个介绍，因为XGBoost在Windows平台上的安装不是那么简单直接。我在实验室的电脑上（Windows 7，64 bits）通过这些步骤安装成功，希望能对后来人有所帮助。

安装必要的软件

为了能在Windows上通过Python使用XGBoost，需要先安装以下三个软件：

- Python
- Git
- MINGW

Python和Git的安装

对于Python，你可以到Python官网上下载你想安装的版本，安装很简单，这里就跳过。对于Git的安装有很多种选择，一种选择就是使用Git for Windows，Git for Windows的安装也比较简单，遵从指示就行，这里也跳过。

XGBoost的下载

Git安装完成后，开始菜单中会出现一个叫Git Bash的程序，点开后就会出现一个类似Windows命令行的窗口，首先在这个Bash窗口，使用cd命令进入你想保存XGBoost代码的文件夹，比如下面的示例：

© 2016 - 2017 ❤ Wang Shuo

然后输入下面的代码下载XGBoost文件包：

```
1 $ git clone --recursive https://github.com/dmlc/xgboost
2 $ cd xgboost
3 $ git submodule init
4 $ git submodule update
```

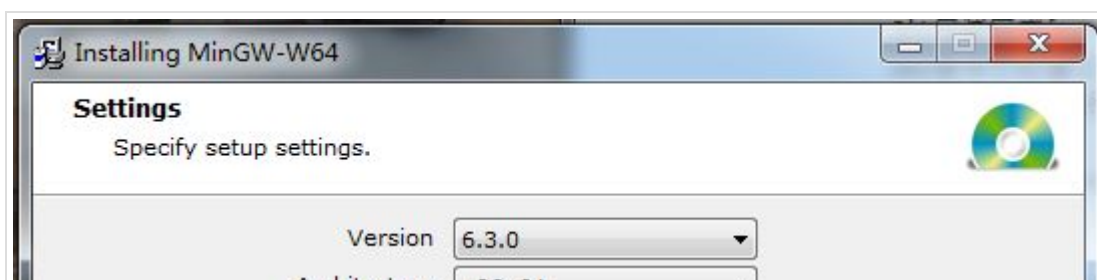
编译XGBoost代码

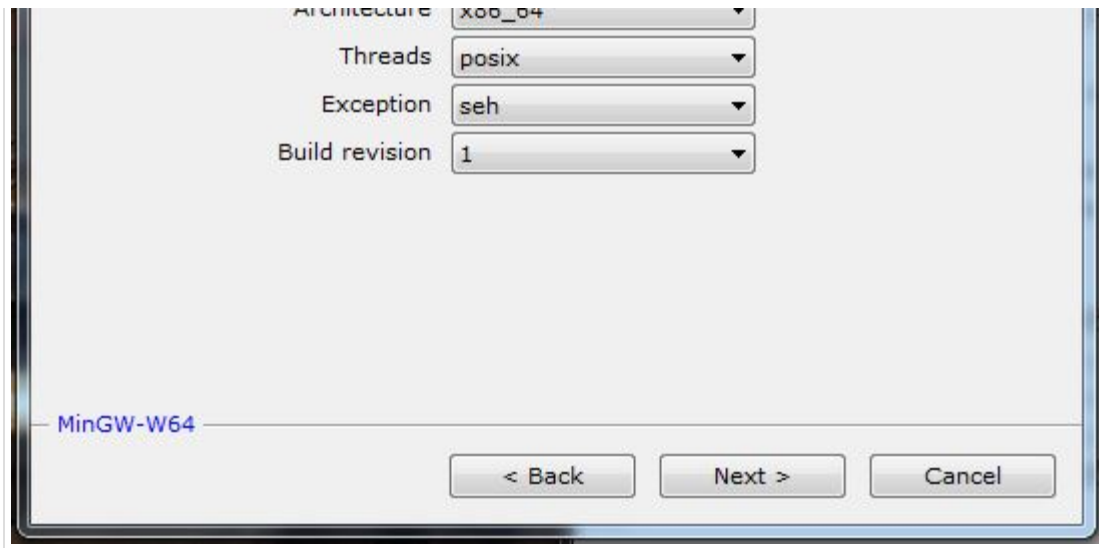
MinGW-W64的安装

接下来就是编译我们刚刚下载的XGBoost的代码。这就需要用到MinGW-W64。它的安装包我是从[这里](#)下载的，下载完成后双击安装，出现下面的安装界面，点击Next：



然后在Architecture选项处选择x86_64即可，其他选项保持默认，如下图：





然后点击下一步，就能安装完成。我使用的是默认安装路径C:\Program Files\mingw-w64\x86_64-6.3.0-posix-seh-rt_v5-rev1。那么make命令和运行库就在下面的文件夹中（也就是包含mingw32-make的文件夹）：C:\Program Files\mingw-w64\x86_64-6.3.0-posix-seh-rt_v5-rev1\mingw64\bin，接下来就是把上面的路径添加到系统的Path中，关于如何添加环境变量到系统的Path中，可以参考[这篇文章](#)。

上面的步骤完成后，关闭Git Bash窗口后重新打开，为了确认添加环境变量已经添加成功，可以在Bash中键入下面的命令：

```
1 $ which mingw32-make
```

如果添加成功的话，应该返回类似下面这样的信息：

```
1 C:\Program Files\mingw-w64\x86_64-6.3.0-posix-seh-rt_v5-rev1\mingw64\bin\mingw32-r
```

为了输入的方便，可以简化mingw32-make命令为make

```
1 $ alias make='mingw32-make'
```

XGBoost的编译

现在就可以开始编译XGBoost了，首先进入xgboost文件夹

```
1 $ cd /e/algorithm/xgboost
```

通过这篇[官方文档](#)给出的统一编译的方法在写这篇文章时还不能正常编译成功，所以我们采用下面的命令来分开编译，每次编译一个子模块。注意，我们要等每个命令编译完成后才能键入下一个命令。

```
1 $ cd dmlc-core
2 $ make -j4
3 $ cd ../rabit
4 $ make lib/librabit_empty.a -j4
5 $ cd ..
6 $ cp make/mingw64.mk config.mk
7 $ make -j4
```

一旦最后一个命令完成后，整个编译过程就完成了。下面就开始安装Python模块。进入XGBoost文件夹下面的python-package子文件夹，然后键入：

```
1 $ cd /e/algorithm/xgboost/python-package>python setup.py install
```

进行到这儿，基本上就完成了，这时打开一个Jupyter notebook，直接导入xgboost包会出现错误，我们需要先运行下面的代码：

```
1 import os
2
3 mingw_path = 'C:\Program Files\mingw-w64\x86_64-6.3.0-posix-seh-rt_v5-rev1\mingw64'
4
5 os.environ['PATH'] = mingw_path + ';' + os.environ['PATH']
```

成功示例

然后我们就可以开始导入xgboost包去运行下面的示例：

```
1 import xgboost as xgb
2 import numpy as np
3
4 data = np.random.rand(5,10) # 5 entities, each contains 10 features
```

```
5 label = np.random.randint(2, size=5) # binary target
6 dtrain = xgb.DMatrix( data, label=label)
7
8 dtest = dtrain
9
10 param = {'bst:max_depth':2, 'bst:eta':1, 'silent':1, 'objective':'binary:logistic'
11 param['nthread'] = 4
12 param['eval_metric'] = 'auc'
13
14 evallist = [(dtest,'eval'), (dtrain,'train')]
15
16 num_round = 10
17 bst = xgb.train( param, dtrain, num_round, evallist )
18
19 bst.dump_model('dump.raw.txt')
```

至此，如果没有出现错误，就表示安装成功。

参考

Jean Francois Puget: [Installing XGBoost for Anaconda on Windows](#)

[#tutorial](#)

◀ 什么是费曼技巧

Resilio Sync : 文件共享神器 ▶