



验证码破解技术四部曲之环境搭建篇（一）

Jun 28, 2016

前言

转眼就要变成大四狗了，大学期间做的比较深入的技术是爬虫，但是爬虫也有很多高级的技术没有涉及，比如说验证码的破解便是其中之一，再加上我对其非常感兴趣，于是乎，开始苦学图像处理、学习机器学习，验证码破解也有了些眉头。

以下是我破解的几种验证码。

1、最容易破解的验证码

直接使用tesseract库识别。

2、download.csdn的验证码

使用K近邻算法识别。

3、weibo.cn验证码

使用卷积神经网络训练识别。

说明

项目源码：<https://github.com/nladuo/captcha-break> (可以帮我点个star(^__^))

开发语言：python(编写爬虫),c++(编写图像处理部分以及机器学习算法)

开发环境：ubuntu 14.04

151.101.36.133

依赖库：

Python：PIL、BeautifulSoup4、requests

C++：boost、opencv2、tesseract-ocr

环境搭建

安装opencv 2.4.12

```
sudo apt-get install build-essential cmake libgtk2.0-dev pkg-config python-dev python-numpy lib
wget https://code.load.github.com/Itseez/opencv/zip/2.4.12
unzip opencv-2.4.12.zip
cd opencv-2.4.12 ; mkdir release ; cd release
cmake -D CMAKE_BUILD_TYPE=RELEASE -D CMAKE_INSTALL_PREFIX=/usr/local ..
make -j8
sudo make install
```

安装tesseract-ocr

```
sudo apt-get install tesseract-ocr tesseract-ocr-dev
```

安装boost

```
sudo apt-get install libboost-dev
```

安装python的PIL库

```
sudo apt-get install python-imaging
```

安装python的BeautifulSoup4、requests库

```
sudo apt-get install python-pip
sudo pip install bs4
sudo pip install requests
```

下载项目源码

```
git clone https://github.com/nladuo/captcha-break.git
```

社交帐号登录: 微信 微博 QQ 人人 [更多»](#)



说点什么吧...

发布

0条评论

[最新](#) [最早](#) [最热](#)

还没有评论，沙发等你来抢

True's ME正在使用多说