

# anti-hack11

---

个人认为这道题是misc中最新颖、神奇的一道题。题目给了一个网页，要求在一定时间内输入100个SJTU验证码，而且最多只能错5次。还给了一个.py文件，看起来是用来分割验证码的。

(看到这道题之后，我觉得别的验证码是用来检测你是不是人类的，这个验证码是用来检测你是不是机器或者超人类的，显然不可能靠手速取胜，那么就要使用机器来识别验证码了。)

问题就在于，题目中给的ASS AI并不总是那么聪明，它总可能会识别出错，但是这意味着我也得跟着错。于是，我们尽可能保持和题目一样的环境，使用它提供的分割验证码的代码进行验证码的分割。

识别验证码，当然首先考虑使用机器学习的方法了。首先，训练使用的数据是网站本身就能提供的（只要每次提交空的内容，网站就会自动显示它的识别结果）。因此，使用Selenium编写一个爬虫，用来爬取验证码数据。

数据获取完成之后，使用PyTorch搭建一个简单的CNN网络进行训练。我的数据样本大概是3000，训练了上百个EPOCH之后就可以得到不错的识别结果了。

之后，再次使用Selenium写爬虫提交验证码，发现识别的准确率并没有想象中的高。每次大概能够识别到70之后，错误次数就超了。于是，我使用了一个小trick: 在识别达到85次之后，先“人工验证”识别结果之后，再提交，确保万无一失。

最终，使用这种方法，得到了flag。代码很长，就不直接贴在这里了，文末附上代码链接。

## 链接

- [爬虫](#)
- [图片分割](#)
- [神经网络](#)
- [提交](#)