# 文本分类

语料：搜狗新闻语料迷你版

运行环境：粒子云远程开发环境

我们使用了3种工具进行文本分类，分别是:

* Sklearn下的Multinomail Naive Bayes
* fastNLP 配合DPCNN 神经网络
* hanlp 下的 Naïve Bayes Classifier

结果概览

|  |  |  |  |
| --- | --- | --- | --- |
|  | 精确度 | 运行时间(总) | 分类时间（不含训练） |
| Sklearn | 0.9207467418104968 | 148.8s |  |
| fastNLP |  |  |  |
| hanlp |  |  |  |

运行时间（总）包含：读取语料，预处理，模型构建与训练，精确度测试。

精确度取第1次运行获得的精确度

精确度（10）取10次测试结果的精确度的平均值