**人工智能文本分类实验验收**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 序号 | 学号+姓名 | 验收项目 | | | | | | 评分 |
| 训练集大小 | 测试集大小 | 字典处理 | 特殊处理 | 选择的分类器 | 性能 |
| 57 | ➀学号： 2018211290  姓名： 崔思颖  分工：数据处理、tf idf计算  ➁学号： 2018211251  姓名： 黄芝遵  分工： SVM分类器训练及测试  ➂学号： 2018211507  姓名： 陈姝仪  分工： 朴素贝叶斯分类器训练及测试 | 共 10 类，  每类 5000篇  □ 自编爬虫  □ 自行爬取  √ 已有语库  □ 混合 | 共 10 类，  每类 5000篇  □ 自编爬虫  □ 自行爬取  √ 已有语库  □ 混合  √训练集/测试集之比:1 | √ 降维  维度 2000  方法tf/idf  √ 停用词是否合理问题： | √ 改进加权  方法： tf/idf  □ 其他改进  方法：  √ 是否有混淆矩阵 | √ 自编贝叶斯  √ 零概率处理:  m-估计 | ◎最高的正确率=98.2323%  ◎最高的召回率=97.6200%  ◎最低的正确率=85.9300%  ◎最低的召回率=86.9600%  ◎平均正确率=93.2956%  ◎平均召回率=93.0620%  ◎训练时间=55s  ◎测试时间=1min40s |  |
| √ SVM（选作）  √ libsvm  ○ 其他  √ 核 RBF  √ 调参 C:35.0 g:0.0079125  ○ 交叉验证  □ 其他分类器 | ◎最高的正确率=97.0291%  ◎最高的召回率=97.7800%  ◎最低的正确率=90.1973%  ◎最低的召回率=90.5400%  ◎平均正确率=94.3370%  ◎平均召回率=94.3240%  ◎训练时间=19min  ◎测试时间=7min |  |

评语：