

数据挖掘大作业程序使用说明

胡玮玮(2006011293) 孔祥欣(2006011299) 徐恺(2006011281)

清华大学计算机科学与技术系计 62 班

{huww06, kxx006, jaguarxk000}@gmail.com

执行文件

在提交的包中，bin 文件夹下有批处理文件 classifier.bat 和 jar 文件 classifier.jar。

使用说明

- 1、双击批处理文件 classifier.bat，此处一定注意是通过该批处理文件启动而不是直接使用 classifier.jar，因为该程序使用到的数据文件比较大，已经超过了 java 虚拟机的大小，所以需要使用额外的命令来增大 java 虚拟机的大小，而这个被写在了批处理文件中。
- 2、启动以后出现如下界面：



- 3、界面介绍

可以看到界面主体分为两个部分，左边是数据文件输入以及参数设置部分，右面是结果显示以及开始按钮部分。左边最上面是用来选择训练集文件，中间可以用来选择测试方式，下面用来进行特征选择算法和分类算法的参数设置。

4、典型的操作过程

启动系统以后，首先输入训练集两个文件，通过“选择特征文件”按钮选择文件选择器默认路径下的 `ad.names` 文件。通过“选择数据文件”按钮选择文件选择器默认路径下的 `ad.data` 文件。

然后可以有两种方式对分类器进行测试，即“来自训练集”和“来自文件”，如果选择“来自训练集”则将对训练集的数据使用 `leave one out` 方法进行循环测试，如果选择“来自文件”则可以选择文件选择器默认路径下的 `ad-test.arff` 文件进行测试。

进而选择特征选择方法和分类算法，并以此对各自的参数进行设置，相关参数的意义请参考实验报告。

所有的左边部分都设置完全以后，点击右侧上方“开始”按钮，则系统自动完成数据挖掘的整个过程，并在右侧下方文本框内显示分类结果，如果在分类过程中你希望停止分类，则点击“停止”按钮即可。

注意事项

由于时间关系，该系统并没有进行太多的容错处理，所以在输入文件的时候，务必确保输入的文件是正确的，否则可能导致不可预知的程序崩溃错误。