

J48又称C4.5，是用在机器学习和数据挖掘的分类问题中的算法，它的目标是监督学习。一个数据集，其中的每一个元组都能用一组属性值来描述。C4.5的目标是通过学习，找到一个从属性值到类别的映射关系，并且这个映射能用于对新的类别未知的实体进行分类。此方法是利用决策树结构来实现的。

1. `evaluation.evaluateModelOnceAndRecordPrediction(j48,testInstance)`: 进行正确率的计算，确定预测是否正确
2. `setClassIndex(data.numAttributes() - 1)`: 为数据集设置类索引，让分类器知道哪个是结果变量
3. `j48.buildClassifier(data)`: 使用j48进行分析数据

因为没有找到要求数据集的arff文件，只找到了.data文件
所以在处理数据前需要先转化文件

通过ConverterUtils.DataSource.read()读取csv

再通过ArffSaver保存arff文件

通过创建FileReader对象，并让其读取对应的arff文件。

通过此Reader对象, 创建Instances对象

此后，`setClassIndex()`为数据集设置类索引，让分类器知道哪个是结果变量

通过J48中的buildClassifier进行分析数据

最后通过Evaluation中的方法进行正确率的计算

三、实验结果

