Weka[4] 特征选择

作者：Koala++/屈伟

特征选择，我对这一部分也不熟，大概讲一下，用 AttributeSelection 进行特征选择，它需要设置 3 个方面，第一：对属性评价的类（自己到 Weka 软件里看一下，英文 Attribute Evaluator），第二：搜索的方式（自己到 Weka 软件里看一下，英文 Search Method），第三：就是你要进行特征选择的数据集了。最后调用 Filter 的静态方法 userFilter，感觉写的都是废话，一看代码就明白了。唯一值得一说的也就是别把 AttributeSelection 的包加错了，代码旁边有注释。

另一个函数懒的解释了（它也不是我写的），基本上是自解释的，不太可能看不懂。

```
package instanceTest;

import java.io.FileReader;
import java.util.Random;

import weka.attributeSelection.CfsSubsetEval;
import weka.attributeSelection.GreedyStepwise;
import weka.classifiers.Evaluation;
import weka.classifiers.meta.AttributeSelectedClassifier;
import weka.classifiers.trees.J48;
import weka.core.Instances;
import weka.filters.Filter;
import weka.filters.supervised.attribute.AttributeSelection;

public class FilterTest
{
    private Instances m_instances = null;

    public void getFileInstances( String fileName ) throws Exception
    {
        FileReader frData = new FileReader( fileName );
        m_instances = new Instances( frData );

        m_instances.setClassIndex( m_instances.numAttributes() - 1 );
    }

    public void selectAttUseFilter() throws Exception
    {
        AttributeSelection filter = new AttributeSelection();  // package
weka.filters.supervised.attribute!
        CfsSubsetEval eval = new CfsSubsetEval();
        GreedyStepwise search = new GreedyStepwise();
        filter.setEvaluator(eval);
        filter.setSearch(search);
        filter.setInputFormat( m_instances );

        System.out.println( "number of instance attribute = " +
m_instances.numAttributes() );

        Instances selectedIns = Filter.useFilter( m_instances, filter);
        System.out.println( "number of selected instance attribute = " +
selectedIns.numAttributes() );
```

```java
    }

    public void selectAttUseMC() throws Exception
    {
         AttributeSelectedClassifier classifier = new AttributeSelectedClassifier();
         CfsSubsetEval eval = new CfsSubsetEval();
         GreedyStepwise search = new GreedyStepwise();
         J48 base = new J48();
         classifier.setClassifier( base );
         classifier.setEvaluator( eval );
         classifier.setSearch( search );
         // 10-fold cross-validation
         Evaluation evaluation = new Evaluation( m_instances );
         evaluation.crossValidateModel(classifier, m_instances, 10, new Random(1));
         System.out.println( evaluation.toSummaryString() );
    }

    public static void main( String[] args ) throws Exception
    {
        FilterTest filter = new FilterTest();

        filter.getFileInstances( "F://Program Files//Weka-3-4//data//soybean.arff");
        filter.selectAttUseFilter();

        filter.selectAttUseMC();
    }
}
```