

Weka[1] Instances 类

作者: Koala++/屈伟

首先,我不懂 java,如果高手看到我的代码有错,就不用发贴鄙视了,直接把正确的发送到我的邮箱 quweiprotoss@163.com 中就好了,我会在更新中表示感谢。

言归正传,开始介绍 Weka,先 google 一下,把 Weka 软件下载下来,安装完成之后,在 Weka 的安装目录中有一个 weka.jar 的包。

把包添加到工程中后,就可以调用 weka 中的函数了。

再介绍一点 weka 的基本知识,在 weka 的目录下,有一个 data 的文件夹,里面存放的是一些数据集,以第一个数据集 contact-lenses.arff 为例,用 EditPlus 或是别的编辑器打开数据集,以%开头的表示的是一些注释,@relation 表示这个数据集的名字,@attribute 表示属性的属性,@data 后就是数据集了,一般来说最后一列是类别(在 LibSVM 中第一列是类别)。

下面我写了一个微不足道的一个 weka 的入门代码:

```
package instanceTest;

import java.io.FileReader;
import weka.core.Instances;

public class InstanceTest
{
    public static Instances getFileInstances( String fileName ) throws Exception
    {
        FileReader frData = new FileReader( fileName );
        Instances data = new Instances( frData );

        return data;
    }

    public static void main(String[] args) throws Exception
    {
        Instances instances = getFileInstances( "F://Program
        Files//Weka-3-4//data//contact-lenses.arff" );

        //把数据集全部输入出
        //System.out.println( instances );

        //用numInstances可以获得数据集中有多少样本
        for( int i = 0; i < instances.numInstances(); i++ )
        {
            //instance( i )是得到第i个样本
            System.out.println( instances.instance( i ) );
        }
    }
}
```