

Weka[5] 半监督算法

作者: Koala++/屈伟

这次介绍一个我现在正在做的半监督算法,因为我只是提出一个框架,半监督算法只是一个工具,可是为了找这个工具也让我花了不少功夫。现在介绍的暂时不是 Weka 的一部分,但开发者是 waikato 大学(也就是开发 Weka 的那个大学)。下载地址是 <http://www.cs.waikato.ac.nz/~fracpete/projects/collective-classification/>。我下载的是 **Classifiers compatible now with Weka 3.5.8**。别的几种,能不能行我也不知道,我然后也按作者说的把 Weka3.5.8 也下载下来了(不知道和以前的版本是否兼容),使用非常简单,先到 Weka 目录下,把 weka-src.jar 解压,再把下载的 src.tar.gz 在 Weka 目录下解压,如果有替换就替换,如果没有替换,证明你的操作有错。

然后就可以试一下了,把全部 Weka 源文件拷贝到你新的工程中去,下面我给出了一个测试的例子,不要太相信我的办法,我还没细看他的代码(分成训练和测试样本的方法有问题,懒得改了,谁改完了,请发给我)。注意一下,他训练是同时用训练集和测试集。如果不明白,下载他的一篇文章 Using Weighted Nearest Neighbor to Benefit from Unlabeled Data。非常简单,不用害怕。

```
package semiTest;

import java.io.BufferedWriter;
import java.io.FileReader;
import java.io.FileWriter;
import java.io.IOException;

import weka.classifiers.Evaluation;
import weka.classifiers.collective.functions.LLGC;
import weka.classifiers.collective.meta.CollectiveBagging;
import weka.classifiers.collective.meta.YATSI;
import weka.classifiers.trees.J48;
import weka.core.Instances;
import weka.core.Utils;
import weka.filters.Filter;
import weka.filters.unsupervised.instance.Resample;

public class Test
{
    Instances m_instances = null;
    Instances m_testIns = null;
    Instances m_trainIns = 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 writeToArffFile(String newFilePath,Instances ins) throws IOException
    {
        BufferedWriter writer = new BufferedWriter(new FileWriter(newFilePath));
        writer.write(ins.toString());
        writer.flush();
    }
}
```

```

        writer.close();
    }

    public void FilterRemovePercentageTest() throws Exception
    {
        Resample removePercentage = new Resample();
        String[] options = Utils.splitOptions("-Z 10 -no-replacement");
        removePercentage.setOptions(options);
        removePercentage.setInputFormat( m_instances );
        m_trainIns = Filter.useFilter( m_instances, removePercentage);
        writeToArffFile("TrainData.arff", m_trainIns );

        options = Utils.splitOptions("-Z 90 -no-replacement");
        removePercentage.setOptions(options);
        removePercentage.setInputFormat( m_instances );
        m_testIns = Filter.useFilter( m_instances, removePercentage);
        writeToArffFile("TestData.arff", m_testIns );

        m_trainIns.setClassIndex( m_trainIns.numAttributes() - 1 );
        m_testIns.setClassIndex( m_testIns.numAttributes() - 1 );
    }

    public void LLGCTest() throws Exception
    {
        System.out.println( " *****LLGC***** " );

        LLGC llgc = new LLGC();
        llgc.buildClassifier( m_trainIns, m_testIns );

        Evaluation eval = new Evaluation( m_trainIns );
        eval.evaluateModel( llgc, m_testIns );
        System.out.println( eval.toSummaryString() );
    }

    public void J48Test() throws Exception
    {
        System.out.println( " *****J48***** " );

        J48 j48 = new J48();
        j48.buildClassifier( m_trainIns );

        Evaluation eval = new Evaluation( m_trainIns );
        eval.evaluateModel( j48, m_testIns );
        System.out.println( eval.toSummaryString() );
    }

    public void YATSItest() throws Exception
    {
        System.out.println( " *****YATSI***** " );

        YATSI yatsi = new YATSI();
        yatsi.buildClassifier( m_trainIns, m_testIns );

        Evaluation eval = new Evaluation( m_trainIns );
        eval.evaluateModel( yatsi, m_testIns );
        System.out.println( eval.toSummaryString() );
    }

```

```

public void CollectiveEMTest() throws Exception
{
    System.out.println( " *****EM***** " );

    CollectiveBagging bagging = new CollectiveBagging();
    bagging.buildClassifier( m_trainIns, m_testIns );

    Evaluation eval = new Evaluation( m_trainIns );
    eval.evaluateModel( bagging, m_testIns );
    System.out.println( eval.toSummaryString() );
}

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

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

    percentage.J48Test();
    percentage.YATSItest();
    //percentage.LLGCTest();
}
}

```