weka 使用指南

陈献钊

天津大学

2019年10月17日



Outline

- ① weka 介绍
- ② 环境配置
 - java 环境配置
 - weka 环境配置
 - 安装 eclipse
 - 配置 eclipse 使得可以使用 weka 的 api
- ③ 使用 java 调用 weka api 实现 k-means 算法
- 4 weka 进一步学习



- ① weka 介绍
- ② 环境配置
 - java 环境配置
 - weka 环境配置
 - 安装 eclipse
 - 配置 eclipse 使得可以使用 weka 的 api
- ③ 使用 java 调用 weka api 实现 k-means 算法
- 4 weka 进一步学习



Weka 的全名是怀卡托智能分析环境(Waikato Environment for Knowledge Analysis),是一款免费的,非商业化(与之对应的是 SPSS 公司商业数据挖掘产品—Clementine)的,基于 JAVA 环境下开源的机器学习(machine learning)以及数据挖掘(data mining)软件。它和它的源代码可在其官方网站下载。有趣的是,该软件的缩写 WEKA 也是新西兰独有的一种鸟名(新西兰秧鸡),而 Weka 的主要开发者同时恰好来自新西兰的怀卡托大学(The University of Waikato)。——摘自百度百科



Outline

- 1 weka 介绍
- ② 环境配置
 - java 环境配置
 - weka 环境配置
 - 安装 eclipse
 - 配置 eclipse 使得可以使用 weka 的 api
- ③ 使用 java 调用 weka api 实现 k-means 算法
- 4 weka 进一步学习



因为 weka 是用 java 写的, 因此我们先下载 java 的 jdk, 到官网 https://www.oracle.com/technetwork/java/javase/downloads/index-jsp-138363.html 下载稳定版的 Java SE 8u231 即可, 注意勾选接 受 License 和根据使用的系统选择对应的架构安装包, 一般都是 Windows x64, 需要使用 oracle 的账户, 注册一个就行



java 环境配置

Java SF 8u231

Java SE 8u231 includes important bug fixes. Oracle strongly recommends that all Java SE 8 users upgrade to this release.

Learn more >

- Installation Instructions
- Release Notes
- Oracle License
- · Java SE Licensing Information User Manual
 - · Includes Third Party Licenses
- Certified System Configurations
- Readme Files
 - JDK ReadMe
 - JRF ReadMe









java 环境配置

weka 介绍

Java SE Development Kit 8u231

You must accept the Oracle Technology Network License Agreement for Oracle Java SE to download this software.

Accept License Agreement

Decline License Agreement

Accept License Agreement		
Product / File Description	File Size	Download
Linux ARM 32 Hard Float ABI	72.9 MB	₹jdk-8u231-linux-arm32-vfp-hflt.tar.gz
Linux ARM 64 Hard Float ABI	69.8 MB	₹jdk-8u231-linux-arm64-vfp-hflt.tar.gz
Linux x86	170.93 MB	- -jdk-8u231-linux-i586.rpm
Linux x86	185.75 MB	- -jdk-8u231-linux-i586.tar.gz
Linux x64	170.32 MB	€jdk-8u231-linux-x64.rpm
Linux x64	185.16 MB	- jdk-8u231-linux-x64.tar.gz
Mac OS X x64	253.4 MB	₹jdk-8u231-macosx-x64.dmg
Solaris SPARC 64-bit (SVR4 package)	132.98 MB	- jdk-8u231-solaris-sparcv9.tar.Z - graph of the state of the stat
Solaris SPARC 64-bit	94.16 MB	- jdk-8u231-solaris-sparcv9.tar.gz
Solaris x64 (SVR4 package)	133.73 MB	₹jdk-8u231-solaris-x64.tar.Z
Solaris x64	91.96 MB	€jdk-8u231-solaris-x64.tar.gz
Windows x86	200.22 MB	€jdk-8u231-windows-i586.exe
Windows x64	210.18 MB	₹jdk-8u231-windows-x64.exe

java 环境配置

然后双击安装包一路选下一步就完事了. 最后在命令行试一试输 入 java -version 有没有反应, 成功会输出相应的 java 版本信息. 没有反应就配置一下 java 的环境变量. 百度自己的系统如何修改 环境变量,加入刚安装的 java 的地址.



weka 环境配置

在 weka 官网 https://www.cs.waikato.ac.nz/ml/weka/找到下载 按钮点击, 选择对应系统的合适版本, 因为已经配置了 ire, 因此 可以下载不带 ire 的版本, 如果选了带 ire 的下载包, 那么在最后 会弹个黑框, 关闭掉后进行 ire 的下载就可以了. 一定要记住 weka 的安装地址, 后面使用 java 编写代码调用 weka 的 api 的时 候要用.



X

weka 环境配置





Choose Components

Full

Choose which features of Weka 3.8.3 you want to install.

Check the components you want to install and uncheck the components you don't want to install. Click Next to continue.

Select the type of install:

Or, select the optional components you wish to install:

Associate Files ✓ Install JRE

Description

Position your mouse over a component to see its description.

Space required: 175.8MB









weka 环境配置

weka 介绍







Space available: 244,5GB

weka 环境配置

最后找到 weka 的文件目录, 双击 weka3.8 就能打开 weka3.8 的 GUI 界面点击 explorer, 点击 open file 导入 weka 文件下 data 文 件夹里的数据文件. 就可以进行可视化的数据分析.



安装 eclipse

因为本文重点并不是 weka GUI 界面的使用, 是使用 JAVA 调用 weka 的 api 进行数据的处理, 因此还需要下载一个 ide 进行 java 代码的编写. 这里使用 eclipse.



安装 eclipse

到 eclipse 官网 https://www.eclipse.org/downloads/下载一个 eclipse. 双击安装包, 选则 eclipse for java developers, 填入你安装 iava 包的位置 (如果没有改动就按默认就行), 然后选择下载的位 置即可. 中途弹出的框点击 accept. 最后还有个弹窗, 选择信赖即 可.



配置 eclipse 使得可以使用 weka 的 api

- 1 打开 eclipse, 选择一个空文件夹作为 workbench
- 2 新建一个 java project
- 3 新建一个 package
- 4 在包里新建一个类
- 5 右键 JRE System Library -> Bulid Path -> Configure Build Path
- 6 点击 Add external JARs 选择 weka 文件夹下的 weka.jar
- 7 关闭对话框



Outline

weka 介绍

- 1 weka 介绍
- ② 环境配置
 - java 环境配置
 - weka 环境配置
 - 安装 eclipse
 - 配置 eclipse 使得可以使用 weka 的 api
- ③ 使用 java 调用 weka api 实现 k-means 算法
- 4 weka 进一步学习



weka 进一步学习

weka 介绍

```
参考 https://www.programcreek.com/2014/02/k-means-
clustering-in-java/
首先引入宏包
import java.io.BufferedReader;
import java.io.FileNotFoundException;
import java.io.FileReader;
import weka.clusterers.SimpleKMeans;
import weka.core.Instances;
```



}

写一个静态类用干读数据文件

```
public static BufferedReader
readDataFile(String filename) {
BufferedReader inputReader = null;
try {
inputReader = new BufferedReader(new FileReader(fil
} catch (FileNotFoundException ex) {
System.err.println("File not found: " + filename);
}
return inputReader;
```

在主体函数中

```
SimpleKMeans kmeans = new SimpleKMeans();
//实例化kmeans类
kmeans.setSeed(10);
//设置随机种子
kmeans.setPreserveInstancesOrder(true);
//设置是否保持样本的顺序
kmeans.setNumClusters(3);
//设置聚类的数目
}
```



```
SimpleKMeans kmeans = new SimpleKMeans();
//实例化kmeans类
kmeans.setSeed(10);
//设置随机种子
kmeans.setPreserveInstancesOrder(true);
//设置是否保持样本的顺序
kmeans.setNumClusters(3);
//设置聚类的数目,这里设置3个因为数据集的任务是分类3
}
```



```
BufferedReader datafile = readDataFile
("D:/cxz/Weka-3-8/data/iris.arff");
// 这 里 的 数 据 集 使 用 weka 自 带 的 安 德 森 鸢 尾 花 卉 数 据 集
Instances data = new Instances(datafile):
//调用 weka的 Instances 类初始化数据集
kmeans.buildClusterer(data);
// 对数据进行聚类划分
```



```
int[] assignments = kmeans.getAssignments();
int i=0;
for(int clusterNum : assignments) {
System.out.printf
("Instance %d -> Cluster %d \n", i, clusterNum);
i++;
}
//翰出结果
```

结果展示

<terminated> Weka [Java Application] C:\Program Files\Java\jre1.8.0_231\bin\javaw.exe (2019年10月17日 下午6:54:59)

Instance 45 -> Cluster 1

Instance 46 -> Cluster 1

Instance 47 -> Cluster 1

Instance 48 -> Cluster 1

Instance 49 -> Cluster 1

Instance 50 -> Cluster 0

Instance 51 -> Cluster 0

Instance 52 -> Cluster 0

Instance 53 -> Cluster 0

Instance 54 -> Cluster 0

Instance 55 -> Cluster 0



Outline

- 1 weka 介绍
- ② 环境配置
 - java 环境配置
 - weka 环境配置
 - 安装 eclipse
 - 配置 eclipse 使得可以使用 weka 的 api
- ③ 使用 java 调用 weka api 实现 k-means 算法
- 4 weka 进一步学习



weka 进一步学习

- I weka 的 api 文档 (英文) http://weka.sourceforge.net/doc.stable/
- 2 weka 的基本使用—explorer 界面的介绍: https://blog.csdn.net/u010372981/article/details/44900435
- 3 WEKA 使用教程: https://blog.csdn.net/yangliuy/article/details/7589306
- 4 官方的 weka 书籍和课程



Thank you for listening!

