

**课程设计报告**

**学生姓名： 高于涵**

**学 号： 19851056**

**学院(系)：计算机与信息技术学院**

**班 级： 计科1904班**

**指导教师： 王志海**

**2022年6月16日**

# 《信息系统集成与开发》课程设计报告

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Java境下开源的机器学习（machine learning）以及数据挖掘（data minining）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New Zealand独有的一种鸟名

## 1 Weka系统简介

数据挖掘、机器学习这些字眼，在一些人看来，是门槛很高的东西。诚然，如果做算法实现甚至算法优化，确实需要很多背景知识。但事实是，绝大多数数据挖掘工程师，不需要去做算法层面的东西。他们的精力，集中在特征提取，算法选择和参数调优上。那么，一个可以方便地提供这些功能的工具，便是十分必要的了。而Weka，便是数据挖掘工具中的佼佼者。[1]

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](http://baike.baidu.com/view/7893.htm)产品--Clementine ）的，基于[JAVA](http://baike.baidu.com/view/29.htm)环Java境下[开源](http://baike.baidu.com/view/9664.htm)的[机器学习](http://baike.baidu.com/view/7956.htm)（machine learning）以及数据挖掘（data minining）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New Zealand独有的一种鸟名，而Weka的主要开发者同时恰好来自New Zealand的the University of Waikato。[3]

Weka作为一个公开的数据挖掘工作平台，集合了大量能承担数据挖掘任务的机器学习算法，包括对数据进行预处理，分类，[回归](http://baike.baidu.com/view/359236.htm)、[聚类](http://baike.baidu.com/view/31801.htm)、[关联规则](http://baike.baidu.com/view/1076817.htm)以及在新的交互式界面上的可视化。而开发者则可使用Java语言，利用Weka的架构上开发出更多的数据挖掘算法。 读者如果想自己实现数据挖掘算法的话，可以看一看Weka的接口文档。在Weka中集成自己的算法甚至借鉴它的方法自己实现可视化工具并不是件很困难的事情。[2]

Weka提供的功能有数据处理，特征选择、分类、回归、聚类、关联规则、可视化等。本文将对Weka的使用做一个简单的介绍，并通过简单的示例，使大家了解使用Weka的流程。本文将仅对图形界面的操作做介绍，不涉及命令行和代码层面的东西。一般常用的有四个应用，分别是：

（1）hanyu汉语（Explorer）

系统提供的最容易使用的图像用户接口。通过选择菜单和填写表单，可以调用Weka的所有功能。这是用来进行数据实验、挖掘的环境，它提供了分类，聚类，关联规则，特征选择，数据可视化等等功能。

（2）Experimenter

用于帮助用户解答实际应用分类和回归技术中遇到的一个基本问题——对于一个已知问题，哪种方法及参数值能够取得最佳效果？通过Weka提供的实验者工作环境，用户可以比较不同的学习方案。尽管探索者界面也能通过交互完成这样的功能，但通过实验者界面，用户可以让处理过程实现自动化。实验者界面更加容易使用不同参数去设置分类器和过滤器，使之运行在一组数据集中，收集性能统计数据，实现重要的测试实验。

（3）KnowledgeFlow

可以使用增量方式的算法来处理大型数据集，用户可以定制处理数据流的方式和顺序。知识流界面允许用户在屏幕上任意拖曳代表学习算法和数据源的图形构件，并以一定的方式和顺序组合在一起。也就是，按照一定顺序将代表数据源、预处理工具、学习算法、评估手段和可视化模块的各构件组合在一起，形成数据流。如果用户选取的过滤器和学习算法具有增量学习功能，那就可以实现大型数据集的增量分批读取和处理。

（4）Simple CLI

是为不提供自己的命令行界面的操作系统提供的，该简单命令行界面用于和用户进行交互，可以直接执行Weka命令。[1]

## 参考文献

[1] 百度百科, URL: https://baike.baidu.com/item/weka/10701215. [2022-6-16].

[2] 取自51CTO博客URL：blog.51cto.com[2022-6-16]

[] 取自程序猿博客URL：http://wjhsh.net/[2022-6-16]