

WEKA

Learning to use a DM tool

by Branko Kavšek

November, 2006

Overview

- What is **WEKA**?
- Preparing the input (data)
- The structure of **WEKA**
- Installing it to your machine
- What else should you know about it?
- Why **WEKA**? / Why not?
- Let's go ... **hands-on** !!!

What is WEKA?

- **WEKA** = **W**aikato **E**nvironment for **K**nowledge **A**nalysis
- An open-source collection of machine learning algorithms in JAVA for data mining tasks
- Contains tools for:
 - data pre-processing
 - classification, regression
 - clustering
 - association rules
 - visualization
- well-suited for developing new ML schemes

What else is WEKA?



Copyright: Martin Kramer (mkramer@wxs.nl)

Preparing the input (data)

- Native format = **ARFF**
- Can manage also:
 - binary serialized instances
 - C45 names files
 - CSV files
 - URLs, databases
 - data generation
- Other formats → CSV → ARFF

The structure of WEKA

- **WEKA** consists of **6** main parts:
 - Simple CLI ~ running WEKA from command-line
 - **Explorer**
 - **Knowledge flow**
 - **Experimenter**
 - ARFF viewer
 - Log window

The installation of WEKA

- JAVA needed
 - WEKA with JAVA
 - WEKA without JAVA
- Different versions of WEKA:
 - stable versions
 - development versions
 - CVS version
- It's a **click-and-go** process !

What else is there to know about WEKA?

WEKA documentation

WEKA homepage:

<http://www.cs.waikato.ac.nz/~ml/weka/>

WEKA book:

Ian H. Witten and Eibe Frank. "***Data Mining: Practical machine learning tools and techniques***", 2nd Edition, Morgan Kaufmann, San Francisco, 2005.

Why WEKA? / Why not?

- It's free
 - It's (relatively) easy to use
 - It's open-source
 - It's well documented
 - Has all the advantages of good OOP
 - It's (almost) all there
- Already have a good implementation of a DM tool that I am used to
 - Need a really fast & efficient algorithm
 - Need a fancy visualization



Let's go ...

... hands-on !!!

