

Lab Tutorial 7

Course: Artificial Intelligence (503030)

I. Objectives

- Overview machine learning.
- Introduction to WEKA tool.

II. Machine Learning

Machine learning is a scientific discipline that explores the construction and study of algorithms that can learn from data. Such algorithms operate by building a model from example inputs and using that to make predictions or decisions, rather than following strictly static program instructions.

For more details, take course CS-503025 – *Machine Learning* (Homepage: <http://it.tdt.edu.vn/~dhphuc/teaching/machine-learning>).

III. WEKA Tool

1. Introduction

WEKA (*Waikato Environment for Knowledge Analysis*) is a popular suite of machine learning software written in Java, developed at the University of Waikato, New Zealand. Weka is free software available under the GNU General Public License.

WEKA is a workbench that contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to this functionality.

Project homepage: <http://www.cs.waikato.ac.nz/ml/index.html>.

WEKA homepage: <http://www.cs.waikato.ac.nz/ml/weka/index.html>.

2. Download and install

Visit: <http://www.cs.waikato.ac.nz/ml/weka/downloading.html>.

3. Manual materials

- [Book] Data Mining: Practical Machine Learning Tools and Techniques - *Ian Witten, Eibe Frank, Mark Hall*.

- WEKA 3.7.12 manual.
 - Keynote of demonstration all graphical user interfaces and how to use Weka for exploratory data mining.
 - ARFF format: <http://weka.wikispaces.com/ARFF>.
- * All materials can obtain at course's homepage, except the book.*

IV. Exercises

1. Introduction the structure of ARFF format.
2. Use the dataset provided on course's homepage to run example of text-classification problem.
The dataset contains:
 - a. ReutersCorn-train.arff: This data set contains 1554 articles, 45 of which are about corn.
 - b. ReutersCorn-test.arff: This data set contains 604 articles for testing purpose.
 - c. ReutersGrain-train.arff: This data set contains 1554 articles, 103 of which are about grain.
 - d. ReutersGrain-test.arff: This data set contains 604 articles for testing purpose.

--- THE END ---