MIDDLE EAST

TECHNICAL UNIVERSITY

COMPUTER ENGINEERING

DEPARTMENT

WEKAREL

User Manual

- Myrzabek MURATALIEV – *1408665*
- Serdar DALGIÇ – *1448570*
- Haşim TİMURTAŞ – *1449149*
- Barış YÜKSEL – *1449826*

Table Of Contents

1. INTRODUCTION....4
1.1. What is Wekarel....4
1.2. Scope of the Document....4
2. Installation....4
2.1. Dependencies....4
2.2. Installation....5
3. HOW TO USE WEKAREL....5
3.1. The Directory Layout....5
3.2. Sample PBS Script....6
3.3. Sample Runs....6
4. References....7

1. INTRODUCTION

1.1. What is Wekarel

Wekarel is the parallel implementation of popular machine learning software Weka, and focuses on achieving data and task parallelism while evaluating clusters. Wekarel, compatible with 3.4 version of Weka, is first attempt when compared with any conjugate software based on Weka.

1.2. Scope of the Document

This document briefly describes the user basic information on how to use Wekarel and get acquainted with the software. It also covers the explanation and examples of the program.

2. Installation

2.1. Dependencies

- Java 1.5 (or later)
- PJ – Alan Kaminsky's Parallel Java Library.
  - Executable version can be downloaded from http://www.cs.rit.edu/~ark/pj.shtml#download
  - The Version (05.04.2009) of pj.jar can also be downloaded from scicomp website: http://senior.ceng.metu.edu.tr/2009/scicomp/docs/pj20090405.jar

2.2. *Installation*

Just download the package and unzip it. After unzipping, place the pj.jar library under *lib* directory. Make sure that pj.jar library is defined in your Classpath.

The program is distributed with the source code inside the package. You can easily browse the source code and contribute to the program in terms of GPL licenses.

3. HOW TO USE WEKAREL

Wekarel has been developed and tested on METU Computer Engineering Department's High Performance Computing Platform – NAR. Selected examples and explanation of the package is given due to this fact. For further implementations and variations of the software, bug – fixes, improvement demands and other stuff, please conduct with the developers via google group : scicomp@googlegroups.com .

The usage of Wekarel is closely related to the ancestor program – Weka. Users are expected to have used or at least tried Weka and have a general knowledge of Data Mining concepts and algorithms.

3.1. *The Directory Layout*

After you unzipped the package, you will see a directory like this:

build build.xml **dist lib src**

build : The executable files of Wekarel built on NAR with Red Hat Linux Distribution.

build.xml: The build.xml file needed to compile the package via Ant. Adapted from original Weka svn code. Not used if source code is compiled by Eclipse or Netbeans.

dist: The distributable package directory.

lib: Library directory for needed libraries. pj.jar file should be included in this

directory and The directory should be introduced in Java CLASSPATH.

src: All the source codes are kept in this directory.

3.2. *Sample PBS Script*

To run weka samples on NAR, a PBS script is needed. Here is a sample PBS Script:

```
#!/bin/sh
#
#PBS -q cengb
#PBS -N SimpleKMeans
#PBS -l nodes=1:ppn=8
#

cd $PBS_O_WORKDIR

pjrun.sh weka.clusterers.SimpleKMeans -t <filename> -N 5
```

pjrun.sh is a sample script located under " /home1/software/ceng490/ " directory. It is written by our Teaching Assistant Celebi Kocair, and aids using PJ library on clusters with Torque (PBS) queue manager. The script is also available on group's download page: http://senior.ceng.metu.edu.tr/2009/scicomp/docs/pjrun.sh

In this example, sample SimpleKMeans parallelization is sent to the Torque queue manager. Number of nodes and cores are specified in the script. For the filename, according to Weka specifications, arff files are used.

3.3. *Sample Runs*

After adequate PBS script is written, *qsub <name of the PBS script>* command is given from the command line. The standard output is written on

<*jobname*>.o<*jobid*> and the standard error on <*jobname*>.e<*jobid*>.

4. References

1. *SciComp Website :* http://senior.ceng.metu.edu.tr/2009/scicomp
2. *Weka Website :* http://www.cs.waikato.ac.nz/ml/weka/
3. *Alan Kaminsky's PJ Parallel Java Library Website :* http://www.cs.rit.edu/~ark/pj.shtml
4. *High Performance Computing Web page of METU Computer Engineering Department :* http://hpc.ceng.metu.edu.tr/