

Data Management

Lab 02 - Analyse de données

Robin Chappatte

Frédéric Montet

Brian Nydegger

Rendu le 25 novembre 2016

à Lausanne

Professeurs :

Dr. Laura Elena Raileanu

Fabien Dutoit

Table des matières

Introduction	1
1 Installation et configuration	2
1.1 Installation	2
1.2 Configuration	2
2 Analyse de données	4
2.1 Arbre de décisions	4
2.2 Clustering	4
2.3 Règles d'associations	4
3 Conclusion	5

Introduction

Pour le second laboratoire du cours de Data Management, nous allons mettre en pratique les concepts d'analyse de donnée vus en cours :

- Arbre de décisions
- Clustering
- Règles d'associations

Pour faire cela, nous utiliserons le logiciel Weka car il dispose d'une grande collection d'outils pour l'analyse de donnée.

1 | Installation et configuration

Dans notre situation nous avons décidé d'installer Weka sur une machine virtuelle Ubuntu 16.10.

1.1 Installation

Nous avons installé :

- Java JDK 1.8_111
- Weka 3.8.0
- Xampp x64 7.0.13
- MySQL connector for Java 5.1.39

Un élément à mentionner est que nous avons utilisé le connecteur dans sa version 5.1.39 alors que sur le site de MySQL¹ il est disponible en 5.1.40. Ceci est dû au fait que nous avons utilisé la version disponible dans un repository ppa. L'installation a pu être faite avec la command `sudo apt-get install libmysql-java`.

Hormis le point cité au paragraphe ci dessus, cette étape n'a pas posé de problème particulier qu'il serait nécessaire de mentionner. Il a suffit de suivre les différents tutoriaux relatifs à chacun des logiciels à installer.

1.2 Configuration

Pour configurer la machine virtuelle, un seul point a posé problème : la connexion à la base de donnée MySQL depuis Weka en utilisant le connecteur MySQL Java.

Le problème rencontré a été la prise en compte de la variable d'environnement `CLASSPATH` lors des tentatives de connexions. Sans cette variable, Weka ne trouve pas le connecteur.

Après avoir effectué une recherche, nous nous sommes rendu compte qu'avec Ubuntu, il fallait préciser le `CLASSPATH` dans la commande pour démarrer Weka à l'aide du paramètre `-cp`. Le listing Figure 1.1 présente le `.sh` qui a été utilisé pour démarrer Weka et le connecter à MySQL avec succès.

1. <https://dev.mysql.com/downloads/connector/j/>

1.2. Configuration

```
1 # Variables
2 java8=/home/fredmontet/Desktop/lab02-weka/asset/java/jdk1.8.0_111/jre/bin/
   ↪ java
3 weka=/home/fredmontet/Desktop/lab02-weka/asset/weka-3-8-0/weka.jar
4 connector=/usr/share/java/mysql-connector-java.jar
5
6 # Command
7 export CLASSPATH=$connector:$CLASSPATH
8 $java8 -Xmx300m -cp "$weka:$connector" weka.gui.GUIChooser
```

FIGURE 1.1 – Contenu du fichier bash à exécuter pour lancer Weka

2 | Analyse de données

2.1 Arbre de décisions

2.2 Clustering

2.3 Règles d'associations

3 | Conclusion