

# note sur weka

Louis LAMALLE

February 2019

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Definition . . . . .	2
1.2	Moyens d'utilisation . . . . .	2
<b>2</b>	<b>Entrainement du modèle</b>	<b>2</b>
2.1	fichier arff . . . . .	2
2.2	Exploitation des données . . . . .	2
2.3	Exemple avec Iris.arff . . . . .	3
<b>3</b>	<b>Utilisation du modèle</b>	<b>3</b>
<b>4</b>	<b>Conclusion</b>	<b>3</b>

# 1 Introduction

## 1.1 Definition

Weka (acronyme pour Waikato environment for knowledge analysis, en français : « environnement Waikato pour l'analyse de connaissances ») est une suite de logiciels d'apprentissage automatique écrite en Java et développée à l'université de Waikato en Nouvelle-Zélande. Weka est un logiciel libre disponible sous la Licence publique générale GNU (GPL).

[3]cf ref

Weka est libre et gratuit. Le logiciel est disponible sur tout les OS. Pour les distributions linux , il suffit d'installer le paquet weka et de lancer la commande weka.

## 1.2 Moyens d'utilisation

Weka peut s'utiliser de différentes manières :

- Grâce à une API , on peut utiliser directement les fonctions de Weka en Java, Python ...
- On peut utiliser Weka directement en ligne de commande grâce à son interpréteur.
- On peut utiliser l'interface graphique de Weka

# 2 Entraînement du modèle

## 2.1 fichier arff

Weka ne fonctionne qu'avec des bases de données en fichier arff (attribut relation file format). C'est un type de fichier créé spécialement pour l'utilisation de Weka. Ils se décomposent en 3 parties :

- Le nom de la table  
ex :  
@RELATION MaRelation
- Les différents attributs. Attention , l'ordre de déclarations des attributs définit l'ordre d'apparitions dans les données (l'attribut déclaré en 3ème position se retrouvera à la troisième colonne des datas)  
ex :  
@ATTRIBUTE nom TYPE
- Les datas , correspondant aux différents objets avec leurs attributs.  
ex :  
@data  
att1.1 att1.2 ...  
att2.1 att2.2 ...

[1]plus sur les arff

## 2.2 Exploitation des données

Une fois la base de donnée créée et télécharger , il est possible de l'exploiter de différentes manières :

- Le Preprocess , il permet de filtrer les données passées en entrée. Par exemple , il est possible de filtrer les attributs de type entier pour les transformer en type nominal ou bien de ne garder que les attributs qui nous intéressent.
- La classification (onglet Classify). Cette fonctionnalité permet de lancer différents algorithmes de prédictions sur la base de données. Par exemple un réseaux de neurones pour effectuer une classification ou alors une régression linéaire pour faire une prédiction.
- Utilisation des méthodes de clustering comme les K-plus proches voisins.
- Accès aux apprentissages par règles d'association qui essaient d'identifier toutes les relations importantes entre les variables.
- Choix des variables les plus pertinentes et prometteuses.
- Affichage graphique scatterplot, arbres.

Au total il y a pls d'une centaine d'algorithmes d'apprentissage disponibles sur Weka.  
[2]Cours de lyon2

### 2.3 Exemple avec Iris.arff

Le fichier test Iris.arff est le plus connu des fichiers pour Weka. Il contient une base de données comportant 3 sortes d'iris avec comme attributs le longueur et la hauteur du pétale et du sépale.

- On ouvre donc le fichier iris.arff préalablement télécharger sur internet (très facile à trouver).
  - On se rend ensuite sur l'onglet classify qui permet de classer les données.
  - Dans l'item choose , on sélectionne l'algorithme désiré. Pour l'exemple , on utilisera le multilayer perceptron (une sorte de réseau de neurones).
  - Une fois sélectionner , on peut avoir accès aux paramètres de l'algorithme dans l'onglet à coté de choose. Il est par exemple possible de définir le coeficient d'apprentissage où le nombre de couches cachées.
  - Il suffit par la suite d'appuyer sur start pour que l'algo se lance et classe les données. En retour , nous pouvons observer plusieurs informations comme le pourcentage d'objets bien classé grâce au modèle.
- Pour les autres algorithmes , beaucoup de fichier exemple arff sont disponibles sur internet pour les tester.

## 3 Utilisation du modèle

Il est possible d'enregistrer son modèle apres l'avoir entrainer. Cela permet de l'utiliser pour la prédiction de données. Un tutoriel est disponible au lien suivant :

<https://machinelearningmastery.com/save-machine-learning-model-make-predictions-weka/>

## 4 Conclusion

Pour conclure avec Weka on peut dire que c'est alternative efficace et très pratique pour la création d'un modèle de prédiction par apprentissage. Bien que l'interface graphique soit assez peu intuitive , beaucoup de ressources sont dispoonible sur internet.

## Références

- [1] arff. URL.
- [2] courslyon2. URL.
- [3] Wekawiki. URL.