
Prise en main du logiciel Weka

Weka est un ensemble de classes et d'algorithmes en Java implémentant les principaux algorithmes de data mining. Il est disponible gratuitement à l'adresse www.cs.waikato.ac.nz/ml/weka, dans des versions pour Unix et Windows. Ce logiciel est développé en parallèle avec un livre : Data Mining par I. Witten et E. Frank (éditions Morgan Kaufmann).

Travail 1 : Identification des composants du logiciel Weka

1. Téléchargez Weka et installez le.
2. Lancer le logiciel Weka et identifier ses composants.
3. Identifier l'onglet de Process de préparation des données dans la rubrique Explorer.
4. Les bases de données exemples sont installées dans le répertoire data. Ouvrez la base d'exemples weather (onglet Preprocess - Open File).
5. Combien y a-t-il d'exemples dans la base ?
6. Quels sont les attributs servant à décrire les exemples ? Pour chacun des attributs, quel est son type et les valeurs possibles ?
7. Visualisez la répartition des classes en fonction des valeurs d'attributs (onglet Visualize, attention à modifier la valeur de Jitter et de PointSize pour bien visualiser les exemples).
8. mêmes questions pour la base iris.arff En particulier, que constatez- vous sur l'interaction entre sepalwidth et sepallength ? et entre petal- width et petallength ?
9. Ouvrez le fichier weather.arff grâce à un éditeur de textes. Décrivez l'en-tête et la structure du fichier.
10. On souhaite utiliser un fichier d'exemples qui n'est pas au format arff. Transformez le fichier présent sur la page "<http://archive.ics.uci.edu/ml/datasets/Credit+Approval>" dans un éditeur de textes pour qu'il corresponde au format weka (fichier .arff).
11. Ouvrez le fichier dans weka et vérifiez que tout apparaît normalement.
12. Quels sont les possibilités de remplacement des données manquantes existantes dans Weka.
13. Remplacer les données manquantes par un outil Weka.