

Rapport de Projet en Intelligence Artificielle

Julien Sahli et Gaëtan Bossy

May 20, 2020

1 Tâche 1

Nous avons utilisé *Pandas* pour importer les données d'entraînement en *.csv*, puis nous les avons parsées afin qu'elles correspondent au template nécessaire à notre algorithme *ID3*.

2 Tâche 2

Similairement à la tâche précédente, nous avons utilisé *Pandas* pour importer les données d'entraînement, puis nous les avons parsées, et enfin nous avons prédit leur label en utilisant la fonction *classifie* et l'indexation -1 permettant d'accéder au dernier élément d'une *string*.

3 Tâche 5

Quelques modifications de *ID3*, implémentées dans la classe *ID3_advanced* et *NoeudDeDecision_advanced*, se sont révélées nécessaires pour implémenter cette tâche:

- Nous avons calculé l'entropie de chaque séparation potentielle des données pour chaque attribut plutôt que de simplement choisir un attribut et de séparer complètement pour chaque valeur de cet attribut. Ceci a nécessité une modification des fonctions calculant l'entropie et les différentes probabilités ainsi que l'implémentation de quelques fonctions similaires.
- Nous avons modifié le système qui retire les attributs de la liste des attributs potentiels pour un noeud afin qu'il ne retire un attribut que lorsque l'ensemble des données d'un noeud possède la même valeur pour cet attribut. Pour ce faire nous avons décidé de recalculer la liste des attributs et de leur valeurs potentielles à chaque création d'un nouveau noeud dans notre fonction récurrente. Il serait probablement possible d'avoir une solution légèrement plus efficace à ce problème mais la simplicité de notre solution nous convient.

- Nous avons modifié la création des enfants pour s'assurer que chaque noeud aie zéro ou deux enfants correspondants au valeur en dessus ou en dessous du seuil de séparation du noeud.
- Nous avons adapté le système de classification pour refléter qu'il n'existait plus toujours un enfant avec la valeur exacte de l'attribut.
- Nous avons réglé quelques problèmes causé par la représentation qui était prévue pour des *strings*.

Ces trois modifications effectuées, nous avons procédé similairement aux tâches 1 et 2 afin d'importer les données d'entraînement et de test, excepté que les données sont transformées en *float* au lieu de *string*.