

Iterative Dichotomiser 3(ID3)

DJIEMBOU TIENTCHEU VICTOR NICO¹

¹*Université de Yaoundé I,
Faculté des Sciences,
Département d'Informatique,
Etudiant Master SD*

January 12, 2024



Outline

- 1 Context
- 2 Types d'algorithmes d'arbre de décision
- 3 Exemple de situation
- 4 ID3
- 5 Analyse d'ID3 & idée de parallélisation



Context



Types d'algorithmes d'arbre de décision

ID3

- Ross Quinlan en 1986
- Classification
- Entropie
- Données numériques
- Données manquantes

C4.5

- Ross Quinlan en 1993
- Classification
- Entropie & ratio de gain
- Données numériques
- Données manquantes

CART

- Leo Breiman et al. en 1984
- Classification & Regression
- Gini
- Données numériques
- Données manquantes



Exemple de situation

Exemple	Cheveux	Ailes	Aquatique	Animal
1	Oui	Non	Non	Mammifère
2	Non	Oui	Non	Oiseau
3	Non	Non	Oui	Poisson
4	Oui	Non	Non	Mammifère
5	Non	Oui	Non	Oiseau
6	Non	Non	Oui	Poisson
7	Oui	Oui	Non	Mammifère
8	Non	Oui	Oui	Poisson
9	Oui	Non	Oui	Poisson
10	Non	Oui	Oui	Poisson

Figure: Jeu de données exemple



Figure: Arbre de décision attendu

ID3

Algorithme : ID3 Itératif

Données : D la matrice de données

Debut

1. Créer une pile vide
2. Créer un noeud de décision racine contenant D et ajouter à la pile
3. tant que la pile n'est pas vide faire
 - (a) Retirer le noeud de décision au sommet de la pile
 - (b) Si tous les exemples dans le noeud de décision appartiennent à la même classe, créer une feuille correspondante à la classe et l'associer au noeud de décision
 - (c) Sinon
 - choisir l'attribut A qui a le gain d'information maximal pour le noeud de décision
 - créer un noeud de décision avec comme étiquette A et l'associer au noeud racine
 - créer des noeuds de décision enfants correspondant à chaque valeur possible A_i de l'attribut et contenant un sous ensemble résultat du partitionnement des exemples du tableau présent dans le noeud de décision tels que $A = A_i$
 - ajouter les noeuds de décision enfants à la pile
4. retourner l'arbre construit

Fin



Analyse d'ID3 & idée de parallélisation

Taches

- Calcul de gain $\in O(A \cdot N)$
- Séparation des exemples
 $\in O(A \cdot N)$
- Itérations $\in O(N)$

Idée de parallélisation

- Calcul de gain
 $\in O(\frac{A}{N_{thread}} \cdot N)$
- Séparation des exemples
 $\in O(\frac{A}{N_{thread}} \cdot N)$

Merci de votre attention

Victor DJIEMBOU¹

¹ *Université de Yaoundé I,
Faculté des Sciences,
Département d'Informatique,
Etudiant Master SD*