

**Université du Québec à Chicoutimi**

**Département des sciences appliquées**

**8INF436 – FORAGE DE DONNÉES**

**TP1**

**Présenté à**

*Haïfa Nakouri, Ph. D.*

**Par**

*Gabriel Aubin-Moreau, AUBG27070103*

*2023-02-23*

# Présentation de la méthode

## Principe

Le extremely fast decision tree ou Hoeffding AnyTime Tree(HATT) fonctionne presque de la même manière que le very fast decision tree ou VFDT. Il diffère dans la façon à laquelle il gère la division des nœuds. Le HATT utilise le meilleur nœuds disponible attribut disponible jusqu’à ce qu’il en trouve un meilleur pour le remplacer. Alors que le VFDT cherche à avoir un nœud qui n’aura jamais à être remplacé. Le HATT converge vers un arbre « batch » idéal.

## Objectifs

L’objectif de la méthode est d’offrir un arbre plus performant que le VFDT en termes de pouvoir prédictif.

## Algorithme

Une image contenant texte

Description générée automatiquement

Une image contenant texte

Description générée automatiquement

Une image contenant texte

Description générée automatiquement

L’algorithme commence par n’avoir qu’une seule feuille la racine. Ensuite, il incrémente les statistiques de chaque nœud en envoyant un par un les données. S’il juge que le résultat de l’algorithme serait meilleur en faisant une division du nœud, il y aura alors division. À chaque fois qu’une donnée passe chaque nœud est réévalué pour décidé s’il doit être remplacé.

## Avantages

L’arbre EFDT requiert moins d’exemple pour être effectif que l’arbre VFDT et offre de meilleurs performance en terme d’ « accuracy ». L’arbre EFDT s’adapte plus rapidement au « concept drift », car il change plus rapidement sa composition interne.

## Inconvénients

Certains problèmes performent mieux avec l’algorithme VFDT, mais en général l’algorithme EFDT l’emporte selon l’étude cité en bibliographie. Parfois, l’algorithme change un nœud qui possède déjà des sous-arbres adapté aux nouvelles données et mine temporairement la performance de l’arbre avant que le sous-arbre soit complètement adapté.

# Comparaison de la méthode EFDT avec trois autres méthodes(VFDT, CVFDT, DT)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Accuracy | Precision | Recall | F1 | ROC-AUC | Temps exécution(s) | Taille mémoire(MB) | Feuilles | Nœuds | Profondeur |
| EFDT | 83,4% | 45,8% | 2,8% | 5,2% | 53,6% | 7 | 2,53 | 4 | 7 | 4 |
| VFDT | 83,6% | 58,0% | 2,1% | 4,1% | 53,1% | 4 | 0,44 | 3 | 5 | 3 |
| CVFDT | 83,3% | 46,2% | 2,2% | 3,9% | 53,1% | 5 | 2,14 | 6 | 11 | 5 |
| DT | 77,8% | 38,5% | 39,7% | 41,0% | 63,4% | 1 | N.d. | 1071 | 536 | 22 |

Figue 1.4 Tableau comparatif des 4 algorithmes utilisés