

Stage "Apprentissage et classification monotone"

Stagiaire: Laura Nguyen. Encadrant: Christophe Marsala

Juin-Juillet 2018

1 Description du sujet

Les arbres de décision monotones sont une variante des arbres de décision (classiques) [5]. Ils ont pour but la prise en compte d'une relation graduelle susceptible d'exister entre les valeurs d'attributs et la classe. Ce stage a pour but d'étudier ce modèle d'apprentissage et, de façon plus générale, d'étudier les approches existantes d'apprentissage pour la classification monotone.

2 Travail à réaliser

Le travail à réaliser se basera sur les points suivant :

1. Implémentation et expérimentation de l'algorithme de construction d'arbres monotone donné dans [5]
 - le travail sur les arbres de décision réalisé dans le TME de l'IA&DS pourra servir de base à cette implémentation.
 - dans un premier temps, on se concentrera sur :
 - des arbres binaires (comme en IA&DS)
 - la *rank Shannon discrimination measure* (rsdm), la *rank gini discrimination measure* (rgdm) et la mesure *dsm*(\cdot). Les autres mesures pourront être implémentées par la suite, il faudra donc prévoir dans la conception du programme ces extensions possibles.
 - expérimentation sur différentes bases d'apprentissage (par exemples, trouvées sur le site UCI) et une comparaison entre rsdm-dsm et rgdm-dsm sera faite.
 - dans un second temps (selon le planning), les autres mesures *mindsm*, *maxdsm* et *avgdsm* (notées **-dsm*), et les *rank* mesures associées seront implémentées et comparées expérimentalement les unes aux autres.
2. Étude théorique des propriétés des mesures **-dsm* pour mettre en évidence leurs points communs et leurs différences
3. Recherche bibliographique¹ sur les mesures de sélection de variables pour la classification monotone. On ne se limitera pas aux approches par arbres de décision. Les articles suivants pourront servir de points d'entrée : [3, 4, 6, 8, 1, 7]
4. (selon le temps) État de l'art sur les approches de classification monotone par forêts d'arbres de décision (par exemple, [2])

3 Livrables

Les documents suivants pourront être réalisés :

- Librairie Python pour l'apprentissage pour la classification monotone

1. Les articles cités sont disponibles sur le cloud du LIP6.

- algorithme de construction d'arbres de décision monotone
- *rank* mesures
- Rapport de synthèse sur l'étude expérimentation des différentes approches pour construire des arbres de décision monotone
- Rapport de synthèse sur l'étude théorique des mesures **-dsr*
- État de l'art sur les approches en apprentissage pour la classification monotone

Références

- [1] J.R. Cano and S. García. Training set selection for monotonic ordinal classification. 112 :94–105.
- [2] Sergio González, Francisco Herrera, and Salvador García. Monotonic random forest with an ensemble pruning mechanism based on the degree of monotonicity. 33 :367–388.
- [3] Qing-Hua Hu, Mao-Zu Guo, Da-Ren Yu, and Jin-Fu Liu. Information entropy for ordinal classification. *Science China Information Sciences*, 53 :1188–1200.
- [4] Qinghua Hu, Xunjian Che, Lei Zhang, David Zhang, Maozu Guo, and Daren Yu. Rank entropy-based decision trees for monotonic classification. 24(11) :2052–2064.
- [5] Christophe Marsala and Davide Petturiti. Rank discrimination measures for enforcing monotonicity in decision tree induction. *Information Sciences*, 291 :143–171, January 2015.
- [6] Weiwei Pan, Qinghua Hu, Yanping Song, and Daren Yu. Feature selection for monotonic classification via maximizing monotonic dependency. 7(3) :543–555.
- [7] Shenglei Pei and Qinghua Hu. Partially monotonic decision trees. 424 :104–117.
- [8] Yuhua Qian, Hang Xu, Jiye Liang, Bing Liu, and Jieting Wang. Fusing monotonic decision trees. 27 :2717–2728.