

Sujet de stage d'Ingénieur ou Master 2 : Apprentissage et modélisation à base d'arbres pour la prise de décision multi-critères interprétable

- **Objectif** : Ce stage a pour objectif d'étudier et construire une approche de prise de décision multicritère fondée sur les arbres de décision. L'enjeu est d'établir les connexions théoriques entre ces deux familles de méthodes afin de construire une approche d'apprentissage multicritère pour la prise de décision tirant partie à la fois du potentiel d'interprétabilité et d'efficacité computationnelle des arbres de décision. On travaillera sur un cas concret basé sur des données réelles monotones.

- **Description** : Le sujet du stage s'inscrit dans la perspective d'apprendre un modèle de classification sur des données monotones qui soit à la fois computationnellement efficace et interprétable. Par interprétable, on entend ici une représentation de la règle de classification qui puisse être intelligible par un humain.

On s'intéressera en particulier à une forme spécifique de classification multicritère ordinaire appelée Non-Compensatory Sorting (NCS, cf. [2],[3]). La littérature du problème d'apprentissage d'un modèle NCS à partir d'une *learning set* propose des approches par la programmation mathématique (cf. [7]) et la résolution de problèmes de satisfiabilité booléenne (cf. [1]).

Par ailleurs, on trouve dans la littérature des méthodes pour apprendre des arbres de décision sur des données monotones (e.g. [9], [8]). Les arbres de décisions sont considérés comme de bons modes de représentation du point de vue de l'interprétabilité (cf. Section 10.13 of [5], [6]). Pour autant, il n'est pas toujours facile d'en avoir une compréhension d'ensemble, ce que permet le modèle de classification NCS.

Le sujet du stage consiste à tirer parti du cadre conceptuel et théorique des arbres de décision (cf. CART, [4]) mais aussi de leur potentiel computationnel afin de développer des outils d'apprentissage de modèles NCS en s'appuyant sur le fait que de tels modèles peuvent être décrits comme des arbres de décision ayant des propriétés spécifiques.

Le stagiaire sera chargé d'étudier, d'implémenter et de tester l'approche développée dans le but d'une comparaison avec les performances de méthodes concurrentes dans la littérature.

- **Équipe encadrante** : comprendra des membres du laboratoire MICS, CentraleSupélec et du LIST, CEA : Vincent Mousseau (CentraleSupélec), Wassila Ouerdane (CentraleSupélec), Jean-Philippe Poli (CEA), Myriam Tami (CentraleSupélec).
- **Lieu du stage** : Laboratoire MICS, CentraleSupélec, Université Paris-Saclay, Campus de Gif-sur-Yvette.
- **Durée du stage** : 6 mois, début du stage dès que possible.
- **Financement** : L'indemnité de stage dépendra du profil du candidat. Le stage pourra se poursuivre en doctorat.
- **Profil du candidat** : Le candidat retenu rejoindra le laboratoire MICS de CentraleSupélec. Les candidatures attendues correspondent à des élèves-ingénieurs en dernière année, ou des étudiants en Master 2 Informatique/mathématiques appliquées. Des compétences sont souhaitées dans les domaines du *machine learning* et des modèles multicritères de décision.

- **Personnes à contacter pour plus d'informations et/ou pour la candidature (CV, lettre de motivation, relevés de notes) :**
 - Vincent Mousseau, *email* : `vincent.mousseau@centralesupelec.fr`
 - Wassila Ouerdane, *email* : `wassila.ouerdane@centralesupelec.fr`
 - Jean-Philippe Poli, *email* : `jean-philippe.poli@cea.fr`
 - Myriam Tami, *email* : `myriam.tami@centralesupelec.fr`

References

- [1] K. Belahcène, C. Labreuche, N. Maudet, V. Mousseau, W. Ouerdane, and Y. Chevaleyre, ‘Accountable approval sorting’, in *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI 2018)*, (2018).
- [2] Denis Bouyssou and Thierry Marchant, ‘An axiomatic approach to noncompensatory sorting methods in mcdm, i: The case of two categories’, *European Journal of Operational Research*, **178**(1), 217–245, (2007).
- [3] Denis Bouyssou and Thierry Marchant, ‘An axiomatic approach to noncompensatory sorting methods in mcdm, ii: More than two categories’, *European Journal of Operational Research*, **178**(1), 246–276, (2007).
- [4] Leo Breiman, *Classification and regression trees*, Routledge, 2017.
- [5] Jerome Friedman, Trevor Hastie, and Robert Tibshirani, *The elements of statistical learning*, volume 1, Springer series in statistics New York, 2001.
- [6] Satoshi Hara and Kohei Hayashi, ‘Making tree ensembles interpretable’, *arXiv preprint arXiv:1606.05390*, (2016).
- [7] Agnes Leroy, Vincent Mousseau, and Marc Pirlot, ‘Learning the parameters of a multiple criteria sorting method’, in *Algorithmic Decision Theory*, eds., R. Brafman, F. Roberts, and A. Tsoukiàs, volume 6992 of *Lecture Notes in Artificial Intelligence*, pp. 219–233, (2011). 2nd International Conference on Algorithmic Decision Theory, ADT 2011, Piscataway, NJ, USA.
- [8] Ryan O’Donnell and Rocco A Servedio, ‘Learning monotone decision trees in polynomial time’, *SIAM Journal on Computing*, **37**(3), 827–844, (2007).
- [9] Rob Potharst and Adrianus Johannes Feelders, ‘Classification trees for problems with monotonicity constraints’, *ACM SIGKDD Explorations Newsletter*, **4**(1), 1–10, (2002).