

Stage Master / ingénieur

Parallélisation de l'apprentissage des préférences dans une méthode de classification multicritères

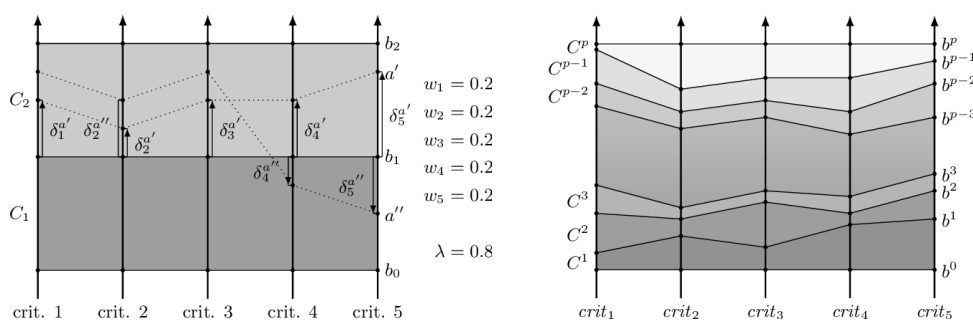
Laboratoire de Mathématique et d'informatique pour la Complexité des Systèmes

MICS — CentraleSupélec — Décembre 2019

1 Contexte

La *preference learning* (PL) est un sous domaine du *machine learning* qui s'intéresse à l'apprentissage des préférences. Les algorithmes de ce domaine sont capables de traiter de grands jeux de données et sont validés au moyen de jeux de données artificiels et réels. Les jeux de données traités en PL sont généralement collectés de différentes sources et sont entachés de bruit. Les algorithmes utilisés dans ce domaine ont des propriétés statistiques fortes leur permettant de s'affranchir du bruit dans ces jeux de données.

Les algorithmes d'apprentissage permettant d'apprendre les paramètres des modèles de classification multicritères nécessitent un nombre de calculs qui peut rapidement exploser avec le nombre de critères (features) et la taille de l'ensemble d'apprentissage (learning set). On considère ici une méthode de classification ordonnée par vote majoritaire (MR-Sort, cf. [Leroy et al. \[2011\]](#), [Sobrie et al. \[2019\]](#)) qui considère des données monotones et fait intervenir une frontière entre classes consécutives (cf. Figure 1). Lors de travaux précédents (cf. [Sobrie et al. \[2019\]](#)), une métaheuristique a été développée permettant d'apprendre les paramètres à partir d'un jeu de données en approximant itérativement le modèle. Afin d'étendre les capacités de cette métaheuristique en termes de taille de l'ensemble d'apprentissage et de nombre de critères (features), il est nécessaire d'utiliser les caractéristiques des architectures modernes telles que le multi-cœur et les GPU.

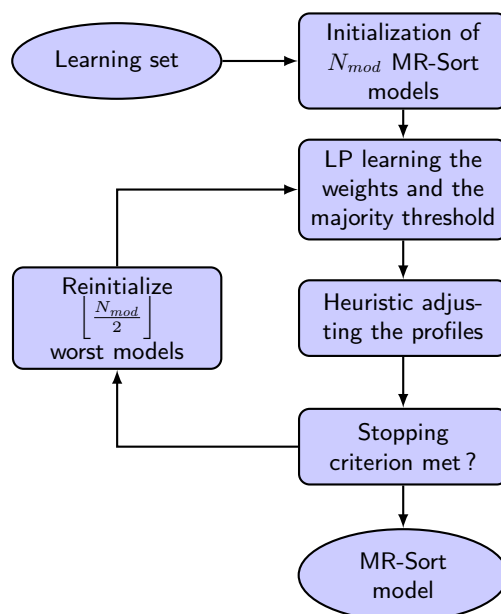


2 Objectifs

L'objectif du stage vise à développer une version parallélisée de la métaheuristique permettant des gains substantiels en termes de temps de calcul, ce qui devrait permettre d'aboutir à la rédaction d'une article scientifique à soumettre dans une conférence/revue internationale. Le travail delandera de :

- Instrumenter la base de code existante pour évaluer conjointement les performances de convergence et de temps d'exécution de chaque partie de l'algorithme.
- Identifier les points clef du processus présentant le plus grand potentiel d'optimisation.

- Proposer et mettre en oeuvre des algorithmes utilisant les capacités de parallélisme des architectures modernes (Multi-cœur, GPU) en explorant les possibilités offertes par Numba, OpenMP ou CUDA selon les solutions envisagées.
- Valider les gains effectifs sur des jeux de données réels (benchmark).



3 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. Des compétences sont souhaitées en modèles de décision/apprentissage/optimisation mais surtout des aptitudes à concevoir/manipuler des algorithmes parallèles sur CPU/GPU.

L'indemnité de stage dépendra du profil du candidat. Le stage pourra se poursuivre en doctorat.

4 Bibliographie

Agnes Leroy, Vincent Mousseau, and Marc Pirlot. Learning the parameters of a multiple criteria sorting method. In R. Brafman, F. Roberts, and A. Tsoukiàs, editors, *Algorithmic Decision Theory*, volume 6992 of *Lecture Notes in Artificial Intelligence*, pages 219–233, 2011. URL http://olivier.sobrie.be/papers/da2pl_2012_sobrie_et_al.pdf. 2nd International Conference on Algorithmic Decision Theory, ADT 2011, Piscataway, NJ, USA.

Olivier Sobrie, Vincent Mousseau, and Marc Pirlot. Learning monotone preferences using a majority rule sorting model. *ITOR*, 26(5) :1786–1809, 2019. doi : 10.1111/itor.12512. URL <https://doi.org/10.1111/itor.12512>.

5 Contacts

Vincent Mousseau - vincent.mousseau@centralesupelec.fr

Laurent Cabaret - laurent.cabaret@centralesupelec.fr