

## SUJET DE STAGE MASTER 2 RECHERCHE

# Algorithmes pour le calcul approché de la constante de normalisation dans les modèles graphiques probabilistes.

**Laboratoire** de Biométrie et Intelligence Artificielle, Institut National de la Recherche Agronomique.

**Lieu:** Auzeville-Tolosane, France

**Projet:** algorithmique approchée pour les modèles graphiques stochastiques (équipes MAD/SaAB)

**Encadrants :** Thomas Schiex & Régis Sabbadin ([Thomas.Schiex@toulouse.inra.fr](mailto:Thomas.Schiex@toulouse.inra.fr) [Regis.Sabbadin@toulouse.inra.fr](mailto:Regis.Sabbadin@toulouse.inra.fr))

**Directeur:** Roger Martin-Clouaire ([rmc@toulouse.inra.fr](mailto:rmc@toulouse.inra.fr))

**Présentation du domaine:** les modèles graphiques probabilistes (réseaux bayésiens, champs de Markov...) permettent une représentation concise de distributions de probabilités sur un produit cartésien de domaines de variables, par un ensemble de fonctions locales, impliquant peu de variables. Ils sont largement utilisés pour la représentation et le raisonnement dans l'incertain en analyse d'image ou en bioinformatique par exemple. Une question centrale dans les modèles graphiques est le calcul de la constante de normalisation (qui permet de transformer un produit de fonctions locales en une distribution de probabilité, normalisée). Ce problème est très difficile ( $\#P$ -complet) et se rencontre lorsque l'on souhaite calculer des lois marginales, par exemple. Il est proche du problème du comptage de solutions dans les problèmes de satisfaction de contraintes (SAT/CSP).

**Objectif:** élaborer une méthode d'approximation de la constante de normalisation dans les modèles graphiques, offrant un compromis réglable entre coût de l'évaluation et qualité de l'approximation. Cette méthode combinera exploration arborescente et exploitation de bornes supérieures et inférieures de la constante de normalisation de sous-problèmes restreints aux variables noninstanciées. Les deux types de bornes qui seront considérées (et qui pourront être combinées et sophistiquées) ont été proposées respectivement (i) dans le domaine des modèles graphiques probabilistes et obtenues analytiquement à partir de l'étude de « distances » entre distribution de probabilités<sup>1</sup>, (ii) dans le domaine des CSP pondérés, avec des bornes obtenues par recherche de cohérence locale pour l'optimisation<sup>2</sup>.

**Travail à réaliser et compétences requises:** dans le cadre de ce stage, l'étudiant devra se familiariser avec le cadre des modèles graphiques probabilistes et celui des CSP pondérés (modèle graphique non stochastique). Il devra en particulier étudier l'extension des bornes obtenues à partir de méthodes de type cohérence locale dans les CSP pondérés pour les étendre au cas de l'intégration discrète et la comparer à la qualité des bornes de type (i). Il devra ensuite combiner des algorithmes d'exploration arborescente « partielle », utilisant ces bornes pour limiter la profondeur des arbres à explorer. Les algorithmes développés seront implémentés (probablement en C++) afin de permettre une comparaison expérimentale sur un ensemble d'exemples liés à la modélisation d'invasion d'espèces nuisibles ou issus d'analyse d'images et de la bioinformatique.

**Indemnité mensuelle:** 417 euros

---

1 Subtree-Based Bounds and simulation-based estimations for the partition function, L. Molokarai. PhD thesis, EPFL, Lausanne, 2007

2 Arc-consistency for soft constraints, MC Cooper, T. Schiex. *Artificial Intelligence*, 2004.