

Optimisation de l'échantillonnage dans les *Markov Logic Networks*

Thématique : Logique / Algorithmique

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

Lieu : Auzeville-Tolosane (près de Toulouse), France.

Equipe : Modélisation des Agro-écosystèmes et Décision.

Projet : Algorithmique pour les modèles graphiques stochastiques.

Encadrants : Nathalie Peyrard et Régis Sabbadin (Nathalie.Peyrard@toulouse.inra.fr,
Regis.Sabbadin@toulouse.inra.fr)

Contexte : le cadre des **Markov Logic Networks** (MLN)¹ fait le pont entre logique du premier ordre (pondérée) et **champs de Markov** (CM, de la famille des modèles graphiques stochastiques). Il offre un grand pouvoir d'expression pour représenter des problèmes de raisonnement ou d'extraction de connaissances. La modélisation en MLN d'une base de connaissances et d'un ensemble d'observations du monde permet, grâce aux outils d'inférence dédiés, d'obtenir de l'information sur les variables non observées. Cependant, ce cadre ne permet pas pour l'instant de traiter le cas où l'on souhaite optimiser le choix des variables que l'on observe. Il s'agit là d'une question d'**échantillonnage optimal**.

Notre équipe possède une certaine expérience autour de l'échantillonnage dans un contexte spatial, à travers le développement d'un cadre basé sur les champs de Markov cachés (CMC)² et de méthodes de conception d'échantillonnage, exactes ou approchées³. Ces outils sont appliqués à la cartographie d'espèces invasives.

Objectif : Dans le cadre du projet ANR « Learning And Reasoning for Deciding Optimally using Numerical and Symbolic information », nous souhaitons maintenant nous appuyer sur nos travaux en échantillonnage spatial pour proposer une **formalisation** et des **algorithmes de résolution** du problème d'échantillonnage dans un MLN. Plus précisément, l'objectif de ce stage est (i) d'étendre le cadre MLN en introduisant dans le langage de représentation des actions d'échantillonnage et des observations et (ii) de proposer des algorithmes approchés d'échantillonnage dans les MLN, s'inspirant de ceux que nous avons déjà définis pour l'échantillonnage dans les CMC.

Travail à réaliser : l'étudiant devra se familiariser avec le cadre des MLN ainsi qu'avec les méthodes d'échantillonnage spatial basées sur les CMC que nous avons développées. Il devra ensuite proposer une extension du cadre MLN prenant en compte des actions d'échantillonnage et des observations, puis proposer une modélisation rigoureuse du problème d'échantillonnage optimal dans les MLN. Des méthodes de résolution approchées devront ensuite être proposées. Nous envisageons une implémentation de ces méthodes via le développement d'une surcouche pour le logiciel *Alchemy*, développé par Domingos et Lowd pour le raisonnement dans les MLN⁴.

Profil recherché : L'étudiant sera issu d'un Master de type « Intelligence Artificielle ». Idéalement, il aura des compétences de base en logique et sera prêt à acquérir des compétences en probabilités (champs de Markov). Par ailleurs, les qualités principales attendues sont la rigueur dans la formalisation et des compétences algorithmiques.

Indemnité mensuelle : 417,09 Euros (M2).

1 Markov Logic : An Interface Layer for Artificial Intelligence. P. Domingos et D. Lowd, Morgan and Claypool, 2009.

2 Decision-Theoretic Optimal Sampling in Hidden Markov Random Fields, N. Peyrard, R. Sabbadin et U. Farrokh, *European Conf on Artificial Intelligence*, Lisbon, 2010.

3 Spatial sampling in HMRF mapping problems : Static and Adaptive algorithms. N. Peyrard et al., *European Conf on Complex Systems*, Lisbon, 2010.

4 <http://alchemy.cs.washington.edu/>