

Échantillonnage de processus spatiaux à partir de connaissances numériques et symboliques

Thématique : Algorithmique / Statistiques.

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

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

Projet : Algorithmique pour les modèles graphiques stochastiques.

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

Contexte : La problématique d'échantillonnage de processus spatiaux consiste à planifier des actions d'observation « locales » d'un processus mal connu, afin de reconstituer une carte de ce processus. Cette problématique relève à la fois des statistiques, pour ce qui est des modèles mathématiques des processus spatiaux (ici les Champs de Markov Cachés, CMC), et de l'intelligence artificielle, pour l'optimisation des stratégies d'échantillonnage (planification dans l'incertain). Elle se rencontre dans des domaines d'application divers : placement de capteurs, diagnostic de systèmes, mais aussi écologie, épidémiologie, gestion de risques de pollution...

Notre équipe possède une certaine expérience de ce problème difficile, à travers le développement d'un cadre pour l'échantillonnage spatial, basé sur les champs de Markov cachés¹ et de méthodes de conception d'échantillonnage, exactes ou approchées² appliquée 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 étendre le cadre et les algorithmes proposés afin de prendre en compte des contraintes « logiques » (à travers l'interdiction de valeurs de configurations) entre les variables cachées, en plus d'informations « numériques » issues de l'échantillonnage. Ces contraintes logiques se révèlent naturellement en intelligence artificielle, où des connaissances symboliques, issues d'experts, sont disponibles, en même temps que des données numériques. Un exemple en écologie est la prise en compte, dans un modèle de cartographie de communautés d'espèces, du fait que certaines espèces ne peuvent pas cohabiter.

Intégrer ces contraintes logiques au cadre des CMC utilisé dans notre approche de l'échantillonnage a des conséquences « mathématiques » (sortie des hypothèses classiques des CMC) et des conséquences sur l'algorithmique d'optimisation de l'échantillonnage. L'examen de ces conséquences fait l'objet du présent stage.

Travail à réaliser : l'étudiant devra se familiariser avec le cadre des CMC ainsi qu'avec les méthodes d'échantillonnage spatial basé sur les CMC que nous avons développées. Il devra ensuite proposer une extension de ce cadre prenant en compte des contraintes logiques, puis examiner l'effet de cette prise en compte (i) sur les algorithmes classiques de restauration des variables cachées dans les CMC et (ii) sur nos algorithmes d'échantillonnage spatial optimal. Ce travail comprend à la fois une partie de formalisation mathématique, et une partie algorithmique. Ce stage pourra comporter une partie de programmation des nouveaux algorithmes développés (étendant des programmes d'échantillonnage spatial existants), si le temps le permet.

Profil recherché : l'étudiant peut être issu d'un Master à dominante mathématique ou d'un Master à dominante informatique. Dans les deux cas, ce stage apportera à l'étudiant une formation complémentaire par rapport à sa discipline d'origine.

Indemnité mensuelle : ~400 Euros (M2).

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

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