

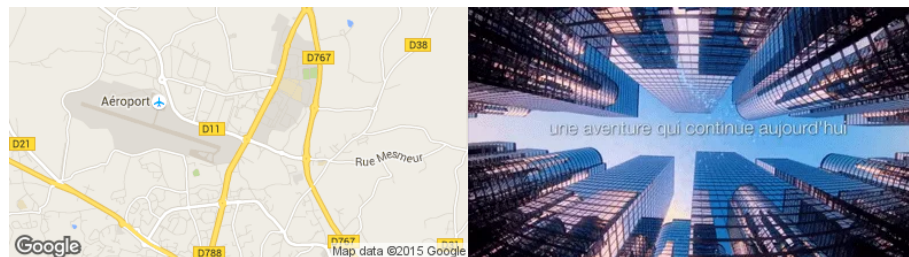
Stage calcul de stratégies optimales et apprentissage par renforcement sur GPU F/H

Mots-clés : calcul, GPU, décision, décisionnel, apprentissage, Orange, Lannion, stage

ref : 0012025 | 02 déc. 2015

date limite de candidature : 30 déc. 2015

2 avenue Pierre Marzin 22300 LANNION - France



Société

Orange



en savoir plus sur nos [étapes de recrutement](#)

Entité

L'équipe d'accueil à Orange Labs, spécialisée en apprentissage statistique et analyse de données, compte actuellement 20 membres permanents et 4 doctorants. Elle est reconnue au sein du groupe pour son expertise. Nous publions régulièrement dans les meilleurs conférences et revues internationales du domaine.

Votre rôle

Sous la direction de Tanguy URVOY, vous contribuerez aux travaux de recherche en apprentissage par renforcement dont les résultats s'ils sont probants pourront faire l'objet d'une publication.

L'avènement de l'internet des objets voit augmenter massivement le nombre de systèmes amenés à prendre de nombreuses microdécisions pour interagir avec un environnement incertain (charge réseau, clients, autres systèmes).

Les processus de décision markoviens (MDP) sont des chaînes de Markov auxquelles on a associé une composante décisionnelle. Ils sont particulièrement adaptés pour modéliser les problèmes de contrôle en environnement incertain (On parle aussi d'apprentissage par renforcement). De tels systèmes doivent simultanément explorer l'espace des états pour acquérir des informations sur leur environnement et exploiter ces informations pour maximiser une fonction de récompense.

Comme expliqué sur le site Wikipédia (https://en.wikipedia.org/wiki/Markov_decision_process) la résolution d'un MDP consiste à calculer une stratégie optimale, c'est-à-dire une fonction qui à chaque état associe une action qui maximise l'espérance de gain à moyen terme du système.

Pour ce faire on utilise des algorithmes de programmation dynamique itératifs comme 'value iteration' (Bellman 1957) ou 'policy iteration' (Howard 1960). Ces algorithmes permettent d'approcher la police optimale en un temps polynomial mais ils restent lourds et difficilement exploitables en pratique.

D'un autre côté, ils sont aussi particulièrement adaptés à un calcul en parallèle.

L'objectif du stage est d'étudier, de coder et d'expérimenter ces algorithmes d'abord dans leur forme séquentielle, puis de manière distribuée sur des processeurs graphiques (GPU Kepler K40) via

l'interface CUDA.

Vous pourrez aussi étudier la méthode des indexes de Gittins qui permet de décomposer un produit de MDPs en une somme de sous-problèmes plus simples.

Votre profil

Vous êtes de niveau Bac +4/5 en informatique, statistique, traitement du signal ou équivalents (écoles d'ingénieurs). Les formations mixtes mathématique/informatique sont bienvenues. Une grande curiosité scientifique, un goût marqué pour la recherche et le travail en équipe, une grande autonomie sont autant de qualités appréciées chez les candidat(e)s.

Ce sujet requiert une double culture en informatique et mathématique appliquée avec quelques notions sur le calcul distribué, l'apprentissage statistique et/ou l'optimisation. Le candidat(e) doit être à l'aise sous UNIX et maîtriser au moins un langage de programmation déclaratif (de préférence le C++ pour les GPUs). Un bon niveau en anglais (écrit et parlé) est nécessaire.

Le plus de l'offre

Le stage se déroulera dans une équipe de haut niveau en recherche et développement, dont l'expertise en traitement statistique des données et apprentissage est reconnue au sein d'Orange et sur le plan international.

La plateforme utilisée pour le stage est récente (GPU Kepler K40).

Les travaux issus de ce stage pourront faire l'objet d'une publication.

Le cadre de vie exceptionnel qu'offre la ville de Lannion.

Contrat

Stage