SFR MathSTIC

"Reinforcement learning"

Séminaire de la SFR MathSTIC

Vendredi 11 octobre 2019, amphi L002, UFR Sciences d'Angers.

~~~

Ce séminaire de la SFR MathSTIC, organisé par Fabien Panloup (LAREMA), David Rousseau (LARIS) et Frédéric Saubion (LERIA), est consacré à l'apprentissage par renforcement, sous divers angles de nos disciplines et à travers différentes applications. Cette après-midi comportera des présentations d'invités ainsi que de membres de nos laboratoires, selon le programme ci-dessous.

12H45 Welcome coffee

Invited talks :

• 13H-13H40 Sylvain Lamprier (LIP6, Paris Sorbonne Université) : Algorithmes de bandits pour la collecte d'informations en temps réel dans les réseaux sociaux.

Nous nous intéressons au problème de la collecte de données en temps réel dans les médias sociaux. En raison des différentes limitations imposées par ces médias, mais aussi de la quantité très importante de leurs données, il n'est pas envisageable de collecter la totalité des données produites par des sites tels que Twitter. Par conséquent, pour être en mesure de récolter des informations pertinentes, relativement à un besoin prédéfini, il est nécessaire de se focaliser sur un sous-ensemble des données existantes. Dans le travail effectué au cours de la thèse de Thibault Gisselbrecht, nous avons considéré chaque utilisateur d'un réseau social comme une source de données pouvant être écoutée à chaque itération d'un processus de collecte, en vue de capturer les données qu'elle produit. Ce processus, dont le but est de maximiser la qualité des informations récoltées, est contraint à chaque pas de temps par le nombre d'utilisateurs pouvant être écoutés simultanément. Le problème de sélection du sous-ensemble de comptes à écouter au fil du temps constitue un problème de décision séquentielle sous contraintes, que nous formalisons comme un problème de bandit avec sélections multiples. Dans cette optique, nous proposons plusieurs modèles visant à identifier en temps réel les utilisateurs les plus pertinents. Dans un premier temps, le cas du bandit dit stochastique, dans lequel chaque utilisateur est associé à une distribution de probabilité stationnaire, a été étudié. Par la suite, nous avons étudié deux modèles de bandit contextuel, l'un stationnaire et l'autre non stationnaire, dans lesquels l'utilité de chaque utilisateur peut être estimée de façon plus efficace en supposant une certaine structure, permettant ainsi de mutualiser l'apprentissage. En particulier, la première approche introduit la notion de profil, qui correspond au comportement moyen de chaque utilisateur. La seconde approche prend en compte l'activité d'un utilisateur à un instant donné pour prédire son comportement futur. Pour finir, nous nous intéressons à des modèles permettant de prendre en compte des dépendances temporelles complexes entre les utilisateurs, grâce à des transitions entre états cachés du système d'une itération à la suivante. Chacune des approches proposées est validée sur des données artificielles et réelles.

• 13H40-14H20 Gilles Fortin-Stoltz (Laboratoire de mathématiques d'Orsay, Université Paris Sud - CNRS) : Pilotage de la consommation électrique par envoi d'incitations tarifaires.

L'électricité se stockant difficilement à grande échelle, l'équilibre entre la consommation et la production d'énergie doit être en permanence maintenu. Actuellement, EDF prévoit la consommation
~~~

électrique et actionne en conséquence ses différents moyens de production. Avec le développement des énergies renouvelables intermittentes sujettes aux changements météorologiques, ajuster la production pour répondre à la demande électrique deviendra de plus en plus complexe. Le déploiement de nouveaux compteurs, capables de collecter les données de consommation mais aussi de communiquer avec les consommateurs quasi-instantanément, permet d'envisager le pilotage de charge.

L'enjeu de ce dernier est de choisir dynamiquement des signaux tarifaires incitatifs à envoyer aux consommateurs afin de moduler leur consommation pour qu'elle s'accorde au mieux avec la production d'électricité : une indication de tarifs plus bas que la référence dans les périodes où l'on voudrait que les clients consomment davantage et anticipent ou reportent autant que possible la consommation qui aurait sinon été effectuée dans des périodes plus tendues (futures ou antérieures) ; ou au contraire, des tarifs plus élevés pour décourager temporairement la consommation.

La difficulté du problème est qu'il faut, simultanément, apprendre la réaction des consommateurs aux différents signaux tout en optimisant l'envoi de ces derniers. Pour bien apprendre les réactions des consommateurs, il faut effectuer une exploration des comportements en jouant sur toute la gamme des signaux, au détriment de l'objectif de pilotage, qui est effectué plutôt en exploitant les résultats de la modélisation des comportements. Nous avons abordé ce dilemme entre exploration et exploitation par la théorie des bandits stochastiques contextuels, et avons été amenés à étendre les résultats théoriques connus au cas où l'on veut suivre une cible de consommation plutôt que maximiser des récompenses, comme c'est classiquement l'objectif.

14H20-14H40 Reinforcement coffee

• 14H40-15H20 Christian Wolf (LIRIS, CNRS, INSA Lyon) : Spatially structured Reinforcement Learning for 3D Control.

In this talk we address the problem of automatically learning the behavoir of intelligent agents navigating in 3D environments from interactions with Deep Reinforcement Learning. We discuss the reasoning capabilities required for these problems on the presence of objects and actors in a scene and to take planification and control decisions. We present a new benchmark and a suite of tasks requiring complex reasoning and exploration in continuous, partially observable 3D environments.

We propose a method, structures its state as a metric map in a bird's eye view, dynamically updated through affine transforms given ego-motion. The semantic meaning of the map's content is not determined before hand or learned from supervision. Instead, projective geometry is used as an inductive bias in deep neural networks. The content of the metric map is learned from interactions and reward, allowing the agent to discover regularities and object affordances from the task itself. We show, that this kind of geometric structure significantly improves the agent's capability of storing objects and their locations and we visualize this reasoning in concrete scenarios.

Local talks :

• 15H20-15H40 : Nicolas Gutowski (ESEO/LERIA) : Recommandation contextuelle de services : Application à la recommandation d'événements culturels dans la ville intelligente.

Cette présentation porte sur les algorithmes de bandits-manchots pour les systèmes de recommandation sensibles au contexte. Les contributions suivantes y sont abordées :
- Diversification des recommandations : Soit via la modification d'un algorithme de bandit-manchot contextuel (Contextual Multi-Armed Bandit: CMAB) : LinUCB ; Soit via une approche porte-folio d'algorithmes de bandits-manchots contextuels et non contextuels (Multi-Armed Bandit: MAB).
- Précision individuelle des recommandations effectuées par les algorithmes de MAB et CMAB.
- Enrichissement dynamique de contexte des algorithmes de CMAB : notamment LinUCB et Contextual Thompson Sampling.
- Application pratique et évaluation en ligne des algorithmes de bandits-manchots pour la recommandation d'événements culturels dans la ville d'Angers : Projet Régional Event-AI (RFI - Atlanstic 2020).

• 15H40-16H00 : Mikael Escobar-Bach (LAREMA, Université d'Angers) : Introduction aux modèles d'urnes et applications à l'apprentissage par renforcement.

Les urnes de Pólya forment un exemple de processus stochastiques par renforcement simples que l'on retrouve dans une grande variété d'applications, allant de la biologie en passant par la finance ou encore l'informatique. Classiquement, on considère une urne remplie de boules de deux couleurs distinctes. A chaque instant, une boule est uniformément tirée au hasard puis remplacée dans l'urne, additionnée d'une autre boule de même couleur et renforçant ainsi le tirage de cette couleur à l'instant suivant.

Dans cette présentation, on se propose donc d'introduire la méthode des urnes accompagnée d'illustrations théoriques et pratiques.

• 16H00-16H20 : Salma Samiei (LARIS, Université d'Angers) : Pedagogical tools to start with reinforcement learning.

In this short talk we will review computational tools and environment read to study or teach reinforcement learning.
