
OFFRE DE STAGE FRANCE-JAPON
MASTER 2 – PFE INGÉNIEUR RECHERCHE

Systemes de recommandation et graphes dynamiques

Contexte scientifique

Le terme de *système de recommandation* regroupe l'ensemble des tâches d'**apprentissage automatique** (*machine learning*) où l'on veut faire correspondre les préférences d'utilisateurs à des articles (contenus culturels, produits, etc.). Les méthodes de l'état-de-l'art reposent typiquement sur des méthodes de factorisation de matrices, reconnues pour leur efficacité.

Récemment, *l'explicabilité* et la *justifiabilité* des systèmes d'apprentissage automatique en général, et des systèmes de recommandations en particulier, ont retenu l'attention de la communauté scientifique. Cependant, ces questionnements se heurtent à de nombreux problèmes; l'un d'entre eux est que l'équité et la diversité ne sont pas formellement définis; un travail fondamental est nécessaire.

Les graphes, c'est-à-dire des ensembles de nœuds reliés entre eux par des arêtes, sont un modèle naturel pour les systèmes de recommandation. On peut recourir à un graphe dirigé, pondéré, dynamique, biparti (ou une combinaison de ces propriétés), selon l'application considérée. Une propriété-clé de ces graphes est leur *interprétabilité*, et en fait des objets très répandus aussi bien en informatique qu'en biologie ou en sociologie.

Le but de ce stage est d'explorer l'apport des graphes et des flots de liens pour les systèmes de recommandation sociaux, après que de premiers travaux aient validé cette approche [3]. Dans un premier temps, il faut définir un ensemble de descripteurs de graphes et de flots de liens adaptés; par exemple, des métriques de similarités telles que l'indice de Jaccard¹, ou encore des métriques de centralité.

Ensuite, nous proposons d'intégrer ces métriques à des algorithmes à l'état de l'art de l'apprentissage automatique : des méthodes classiques, par exemple de *gradient boosting* [1] ou des *factorization machines* [2], ainsi que des méthodes plus récentes d'apprentissage profond autour des réseaux de neurones de graphes.

1. Étant donné deux nœuds u et v et leur voisinages $N(u)$ et $N(v)$, l'indice de Jaccard mesure la similarité entre les deux voisinages : $\mathcal{J}(u, v) = \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|}$.

Déroulement du stage et gratification

Le stage se déroulera entre le laboratoire Cedric du Conservatoire National des Arts et Métiers (3 mois), à Paris, et le Riken Advanced Intelligence Project (3 mois), à Tokyo. La période au Japon devra s'achever au plus tard le 31 août 2019. La période en France pourra être effectuée avant (par exemple, mars-avril-mai), ou après (septembre-octobre-novembre).

Une gratification de stage de 577,05 euros est prévue pour la partie en France, ainsi que la moitié du coût de transport.

Une gratification de stage de 66000 yens (environ 530 euros), ainsi qu'une prise en charge du logement (dans la limite de 100000 yens – environ 850 euros par mois) et la totalité du transport est prévue. Le billet d'avion aller/retour pour Tokyo sera également pris en charge.

Contact

Merci d'envoyer votre demande accompagnée d'un CV aux adresses fournier@cnam.fr et tiphaine.viard@riken.jp.

Mots-clefs

Machine learning – Systèmes de recommandation – Graphes

Références

- [1] J. H. Friedman. Greedy function approximation : a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.
- [2] S. Rendle. Factorization machines. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pages 995–1000. IEEE, 2010.
- [3] T. Viard and R. Fournier-S'niehotta. Augmenting content-based rating prediction with link stream features. *Computer Networks*, 2018.