Offre de Post-Doc « techniques en Machine Learning pour le filtrage, la classification et la prédiction de données pour le management réseau (« SLA enforcement ») en communications multimédia massives » @ Orange Labs (Châtillon, IDF)

Votre Rôle

Les travaux de recherche à mener ont globalement pour but de designer et d'implémenter de nouveaux algorithmes de Machine Learning [1,2] pour le management d'applications réseaux et cloud dans le contexte des futurs réseaux 5G [3].

Pour construire l'architecture d'un système cognitif de management de réseaux, les besoins fonctionnels et non-fonctionnels en termes de services et de business (robustesse, gestion des coûts pour les ressources, Service Level Agreements, responsivité/latence, adaptabilité, durabilité et confiance) doivent être analysés suivant de multiples perspectives. Celles-ci incluent notamment les demandes croissantes pour de meilleures performances et expériences utilisateur pour les réseaux 5G, la large génération de données massives via les réseaux 5G, les préoccupations concernant la sécurité et la privauté des données, les droits de propriété des données (en particulier l'accès aux données depuis différents domaines de régulation).

Les cas d'usage et les scénarii d'intérêt pour Orange concernant le management des réseaux 5G consistent en un certain nombre de services de management d'une infrastructure 5G, tels que :

- les services de monitoring du réseau,
- le service de profiling service (qui sert à automatiser les configurations de l'infrastructure 5G),
- le service de prédiction (pour l'automatisation et la visualisation d'incidents sur le moyen/long terme ou de fraudes dans n'importe quel noeud du réseau),
- le service de provisioning (pour automatiser le déploiement des différentes ressources et les gérer en conformité avec les Service Level Agreements via le réseau 5G),
- le service de gestion d'incident (qui sert à automatiser l'analyse des causes originelles de toute vulnérabilité identifiée dans le réseau 5G).

Plus spécifiquement, le cas d'usage du projet qui sera considéré et adressé par ces travaux de recherche post-doctorale traite de la mise en vigueur des Service Level Agreements dans le contexte de communications multimédia massives. L'état-de-l'art de cette problématique et des solutions associées en termes d'approches issues du Machine Learning est déjà examiné dans un programme d'étude et de recherche doctoral à Orange.

Le but principal de ce programme de recherche post-doctorale proposé est ainsi d'étudier la pertinence et de designer des méthodes d'apprentissage adaptées et optimisées pour des modèles de type réseaux de neurones profonds (c.f. [15,16], parmi d'autres modèles) appliqués à la gestion et à la mise en application de Service Level Agreements dans un contexte d'un scénario de communications multimédia massives, d'intérêt pour Orange.

Dans cette optique, quatre principaux axes de recherche seront adressés:

- Les résultats les plus récents en apprentissage statistique, tels que les méthodes basées sur des noyaux [4,5,6,7] et les espaces de sortie structurés [8], seront considérés afin de modéliser les interactions entre les nœuds du réseau. La théorie des noyaux structurés [8,9,10] sera utilisée pour la modélisation du réseau d'entrée, i. e. la structure de l'ensemble de noeuds, tandis que les techniques de sorties structurées seront utilisées pour modéliser le réseau de sortie, i. e. ils produiront une configuration du réseau (ou à tout le moins une partie de celle-ci).

- Les méthodes d'adaptation de domaine, pour permettre l'utilisation de techniques en apprentissage supervisé [1,2] dans un environnement de type réseau, où la grande variabilité des données nécessite un haut niveau d'abstraction, en sorte que les classifieurs automatiques [1,2] puissent fonctionner efficacement sur des variables d'entrée inconnues (propriétés de généralisation).

- Les approches non-supervisées [1,2] et semi-supervisées [11] afin d'améliorer les capacités d'adaptation de domaine des classifieurs, de façon à ce que ceux-ci puissant fonctionner sous différentes contraintes.

- L'utilisation des méthodes décrites ci-dessus pour implémenter des tâches de prédiction en considérant des modèles de séries temporelles [12] et/ou séquentiels.

En plus des directions de recherche ci-dessus, les activités de recherche aborderont également l'application de bonnes pratiques, traditionnelles et bien évaluées, en Machine Learning à toutes les tâches d'intérêt en classification, pour la régression et pour la prédiction :

- Le traitement de données, qui sont récupérées depuis des sources distinctes, afin de leur donner un format approprié pour être traitées par des algorithmes de Machine Learning. Ces algorithmes apprendront à réaliser plusieurs tâches de classification pour le projet de recherche [3].

- L'ingénierie et l'extraction de features de manière automatique à partir de données brutes. Ces dernières seront transformées au format d'entrée pour les méthodes de Machine Learning, e.g. des représentations basées sur des features, cf. [13].

- Régler les paramètres adéquats pour les différents algorithmes de Machine Learning, appliquer des techniques de sélection automatique de features [13] afin de réduire la dimensionnalité de l'espace de travail ("réduction de la dimension", c.f. [1,2]) et ainsi améliorer l'efficacité des traitements, et filtrer les échantillons d'entraînement qui sont inconsistants.

- La détermination des points forts et des points faibles des différents modèles de Machine Learning, notamment pour assurer une décision de qualité chaque modèle est testé en termes d'efficacité et de précision. Par exemple, nous nous emploierons à appliquer la "n-cross fold validation" [2] pour vérifier si le modèle appris fonctionne de la manière voulue.

Les activités de recherche s'inscriront naturellement au sein du projet de recherche collaboratif EU 5G-PPP "COGNET" [3].

Références

[1] Friedman, J., Hastie, T., & Tibshirani, R. The elements of statistical learning. Springer, Berlin: Springer series in Statistics (Vol. 1), 2001.

[2] Alpaydin, E. Introduction to machine learning. MIT Press, 2014.

[3] EU 5G-PPP « COGNET » project : http://www.cognet.5g-ppp.eu/

[4] Berlinet, A., & Thomas-Agnan, C. Reproducing kernel Hilbert spaces in probability and statistics. Springer Science & Business Media, 2011.

[5] Cristianini, N., & Shawe-Taylor, J. An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, 2000.

[6] Scholkopf, B., & Smola, A. J. Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT Press, 2001.

[7] Shawe-Taylor, J., & Cristianini, N. Kernel methods for pattern analysis. Cambridge University Press, 2004.

[8] Tsochantaridis, I., Hofmann, T., Joachims, T., & Altun, Y. Support vector machine learning for interdependent and structured output spaces. In Proceedings of the 21st International Conference on Machine learning (ICML) 2004.

[9] Yu, C. N. J., & Joachims, T. Learning structural SVMs with latent variables. In Proceedings of the 26th International Conference on Machine Learning (ICML), pp. 1169-1176, 2009.

[10] Vishwanathan, S. V. N., Schraudolph, N. N., Kondor, R., & Borgwardt, K. M. Graph kernels. The Journal of Machine Learning Research, 11, 1201-1242, 2010.

[11] Chapelle, O., Scholkopf, B., & Zien, A. Semi-supervised learning. MIT Press, 2006.

[12] Kantz, H., & Schreiber, T. Nonlinear time series analysis. Cambridge University Press, 2004.

[13] Guyon, I., & Elisseeff, A. An introduction to variable and feature selection. The Journal of Machine Learning Research, 3, 1157-1182, 2003.

[14] Steinwart, I., & Christmann, A. Support vector machines. Springer Science & Business Media, 2008.

[15] LeCun, Y., Bengio, Y., & Hinton, G. Deep learning. Nature, 521(7553), pp. 436-444, 2015.

[16] Goodfellow, I. J., Bengio, Y. & Courville, A. Deep learning. An MIT Press book in preparation, 2016. Draft chapters available in HTML at http://www.deeplearningbook.org/

Votre Profil

- Compétences scientifiques et techniques :

- Mathématiques : modélisation, statistiques et probabilités
- Informatique : maîtrise de Python et/ou LUA et/ou Matlab, ainsi que des bibliothèques courantes implémentant les principaux modèles et méthodes en machine learning
- Des compétences en machine learning/apprentissage statistique sont un atout

- Expériences souhaitées :

- Master et/ou Ecole d'Ingénieur, spécialité mathématiques appliquées et/ou informatique
- Doctorat en mathématiques appliquées et/ou informatique
- Une thèse de doctorat dont les travaux de recherche s'inscrivent dans les thématiques « machine learning » est un atout

Entité

- Département "Modeling and Statistical Analysis" (MSA).
- Activités sur la modélisation des trafics et des réseaux fixes (Internet par Fibre et ADSL) et mobiles (2G/3G/4G →5G).
- Objectifs:
  - Apporter une meilleure connaissance des trafics, en rapport avec l'évolution des usages et des terminaux dans le cadre de la croissance très importante de ces trafics
  - Estimer et optimiser les coûts des infrastructures réseaux par zones géographiques dans le cadre de choix stratégiques de déploiement des technologies de nouvelle génération (fibre optique, réseaux mobiles 4G) avec de nouvelles architectures
  - Améliorer la Qualité de Service (QoS) et les performances des réseaux mobiles
  - Fournir des modèles analytiques de consommation d'énergie des équipements pour estimer et prévoir la consommation d'énergie des réseaux

Durée : 1 an (extensible à 14-15 mois suivant résultats), date de début en 2016 (flexible)

Le Plus de l'Offre

Ce Post-Doc permettra au/à la candidat(e) d'approfondir ses connaissances en télécoms et en mathématiques appliquées et de les mettre en pratique sur un domaine d'application télécom de grande importance.

Il/elle travaillera en collaboration directe avec des ingénieurs de recherche et des chercheurs, tant sur la conception que sur l'implémentation des solutions, et au sein du projet collaboratif EU 5G-PPP COGNET.

Contact

Stéphane Sénécal : stephane.senecal@orange.com