

Stage de recherche

Les réseaux profonds pour les données temporelles multivariées

Application « Jumeau Numériques » sur les moteurs d'avion.

Possibilité de poursuivre en thèse CIFRE avec SAFRAN

Profil candidat : Fin de cycle d'Ingénieur d'une grande école, M2 de data science, statistique et/ou intelligence artificielle. Bonne expérience en programmation, et en particulier du framework PyTorch/deeplearning4j.

Lieux du stage : LIPN, UMR CNRS 7030,

Comment déposer sa candidature : le dossier de candidature en PDF comportera les éléments suivants :

-CV ; Relevés de notes ; Lettre de motivation

Le dossier de candidature est à envoyer par mail à Hanene.Azzag@lipn.univ-paris13.fr, Mustapha.lebbah@univ-Paris13.fr, (objet du mail [Stage-LIPN-SAFRAN-22])

Description du sujet

A travers ce sujet de stage de recherche niveau M2, nous souhaitons tester la viabilité d'une méthodologie neuronale moderne basée sur les architectures profondes (Deep learning), les réseaux récurrents et appliquée aujourd'hui avec succès sur le traitement de textes (traduction, chat-bots, etc.) et l'analyse de signaux audio (sous-titrage automatique). Deux approches sont possibles, une approche anticipative qui identifie un état latent se superposant à l'observation et qui permettrait d'estimer la transition d'une observation à l'autre (de type LSTM ou GRU) [1,2,3]. Une autre proposition est de construire un auto-encodeur récurrent capable de reproduire à l'inverse une série d'observations de manière analogue aux outils utilisés pour l'interprétation des séquences de mots (réseaux *transformers* utilisant des couches d'attention). Ces deux approches nécessitent d'explorer simultanément plusieurs ensembles de séquences multivariées. Il va falloir apprendre à traiter efficacement les signaux temporels multivariés issus des données de vol.

Ce sujet de stage de recherche a pour objectif de mener à un projet CIFRE avec SAFRAN en informatique scientifique et algorithmique qui aura pour but d'intégrer de nouveaux algorithmes sous la méthodologie développée précédemment [4,5].

Le stage de recherche se déroulera en 3 phases :

1. Étudier l'état actuel de l'art sur l'apprentissage profonds et données temporelles multi-variés,
2. Examiner l'état actuel de l'art des outils logiciels et des architectures pour traiter de grande masses de données temporelles multi-variées. Cela comprend les progiciels et les bibliothèques utiles pour construire, former et déployer des modèles sur des données réelles. Cette phase sera réalisée en étroite collaboration avec la Start-up HephIA.
3. Sur la base des études précédentes, implémenter un ou plusieurs algorithmes/architecture. Les résultats obtenus pendant le stage peuvent conduire à des contributions à des logiciels libres, voire à une publication scientifique, en fonction des compétences et de la motivation du/de la stagiaire.

Références

1. Hochreiter, S., Schmidhuber, J.: Long Short-Term Memory. *Neural Computation* 9(8), 1735-1780 (1997)
2. Srivastava, N., Mansimov, E., Salakhutdinov, R.: Unsupervised Learning of Video Representations using LSTMs (2015). <https://doi.org/citeulike-article-id:13519737>, <http://arxiv.org/abs/1502.04681>
3. Madiraju, N.S., Sadat, S.M., Fisher, D., Karimabadi, H.: Deep Temporal Clustering : Fully Unsupervised Learning of Time-Domain Features pp. 111 (2018), <http://arxiv.org/abs/1802.01059>
4. Forest, F., Lebbah, M., Azzag, H., and Lacaille, J. Deep Embedded Self-Organizing Maps for Joint Representation Learning and Topology-Preserving Clustering. *Neural Comput & Applic.* 2021.
5. Forest, F., Cochard, Q., Noyer, C., Joncour, M., Lacaille, J., Lebbah, M., & Azzag, H. (2020). Large-scale Vibration Monitoring of Aircraft Engines from Operational Data using Self-organized Models. *Annual Conference of the PHM Society*, 12(1), 11. <https://www.phmpapers.org/index.php/phmconf/article/view/1131>