

Fiche Stage

CRD/Laboratoire/URD

TECH/ASAP/PROF

Intitulé du Stage

Détection de concept drift

Mission:

L'analyse de flux de données traite des données massives grâce à des algorithmes en ligne qui évitent le stockage exhaustif des données. La détection de changements dans la distribution d'un flux est une question importante dont les applications potentielles sont nombreuses.

En apprentissage supervisé "automatique", on parle de "concept drift" lorsque les propriétés statistiques de la variable cible (Y), que le modèle essaie de prédire, évoluent au cours du temps d'une manière imprévue : $P(Y|X)$ varie au cours du temps.

Le stage sera consacré au développement d'une nouvelle méthode de détection qui sera largement inspirée de l'article "*Détection de changements de distribution dans un flux de données : une approche supervisée - EGC (Extraction et Gestion des connaissances), 25/01/2011, Brest, A. Bondu, M. Boullé*".

Le stage se déroulera en 3 parties : rédaction d'un état de l'art sur le sujet et choix des méthodes concurrentes à tester, rédaction d'un plan d'expérience et choix des bases de données qui serviront aux tests, test de la nouvelle méthode.

Un tutorial sur le sujet : "PAKDD 2011 Handling Concept Drift: Importance, Challenges and Solutions" <http://www.cs.waikato.ac.nz/~abifet/PAKDD2011/>

Une méthode concurrente à tester pendant le stage . "*Segmentation of the mean of heteroscedastic data via cross-validation. Statistics and Computing, 1--20. DOI: 10.1007/s11222-010-9196-x. Sylvain Arlot, Alain Celisse (2010)*"

Profil:

- Le profil souhaité est BAC + 5, Master Recherche ou Master Industriel informatique et/ou statistiques ou école d'ingénieur.
- Intérêt équilibré entre les aspects applicatifs et théoriques du sujet.

Compétences

- Le stagiaire devrait avoir une bonne connaissance des statistiques et des mathématiques.
- Les connaissances en Matlab (ou équivalent) seraient les bienvenues.
- Des connaissances en apprentissage statistique sont un plus.

Modalités

5 mois, printemps-été 2012, Lannion (Bretagne)

Le plus de l'offre

Proche de la mer, vous serez dans l'équipe de traitement des données d'Orange Labs directement en lien avec des problématiques opérationnelles du groupe Orange sur le CRM et l'Audience. Le stagiaire évoluera dans un contexte très recherche sur un sujet très porteur. Il sera intégré dans l'URD au sein d'une équipe recherche.

Contacts

Vincent Lemaire - 02 96 05 31 07 - vincent.lemaire@orange-ftgroup.com

The English-speaking candidacies are accepted