

Stage recherche 2018

Laboratoire d'accueil ULCO : LISIC

Encadrant ULCO : Franck Dufrenois

Collaborations internationales :

- Ira Assent : professeur assistant à l'université d'Arrhus, Danemark
- Gert Cauwenberghs : professeur à l'université de San Diego, Etats Unis

Projet de recherche : Apprentissage actif en ligne pour la détection d'événements anormaux dans les flux vidéo.

Introduction.

Dans de nombreux problèmes d'apprentissage statistiques, il n'est pas rare que les données soient disponibles de manière séquentielle. Il est donc naturel de reformuler le problème de l'apprentissage en tenant compte de cette propriété. Communément dénommé « *apprentissage en ligne* », le classifieur met donc à jour sa frontière de décision de manière *incrémentale*, c'est-à-dire à chaque arrivée d'une nouvelle donnée d'apprentissage. Il se distingue donc des techniques d'apprentissage par lot qui à l'inverse génère une frontière de décision à partir de la totalité des données d'apprentissage. Cependant, une des restrictions à son utilisation est que la collecte des étiquettes des données deviendrait très vite fastidieuse suivant ce mode de fonctionnement car cela nécessiterait de renseigner la classe de la donnée à chaque étape. Une telle démarche trouve donc son intérêt que dans un contexte d'apprentissage *non supervisée*. Les performances de cette approche sont cependant fortement dépendantes de la qualité des données qui lui sont fournies. La présence de données anormales ou inhabituelles dans le flux de données conduit inexorablement à une perte de performance du classifieur. Une stratégie classique consiste à remettre à jour le classifieur lorsque la donnée qui vient d'être classée présente un niveau de confiance faible. Cette incertitude peut être levée en introduisant dans le processus d'apprentissage la connaissance de l'expert qui va permettre de trancher et fournir une étiquette. C'est ce qu'on appelle l'apprentissage *actif* qui conduit à reformuler le problème d'apprentissage en mode semi supervisé.

Le stage de recherche

Le projet de recherche que nous proposons durant ce stage s'inscrit dans ce cadre. Récemment, nous avons proposé une méthode de détection d'anomalies par lot. Lorsque la taille de l'échantillon augmente de manière significative, cette approche est confrontée à des problèmes de stockage et à des problèmes numériques qui la rendent inopérable. Le premier objectif de ce stage est donc de proposer une formulation incrémentale de cette approche afin de réduire l'influence de la taille de l'échantillon. Les travaux sur l'apprentissage incrémental des machines à vecteurs supports proposé par le professeur Gert Cauwenberghs constitueront une base de réflexion pour modifier notre algorithme. Dans un premier temps, cette version incrémentale sera validée sur des jeux d'échantillons synthétiques. Puis, nous l'utiliserons sur données réelles telle que la détection d'événements anormaux (comportement anormal de personnes, présence de bagages suspects dans une gare,...) dans les flux vidéo couleurs. Enfin, nous prolongerons notre étude pour développer une stratégie d'apprentissage active de l'algorithme.

Collaboration internationale

Cette demande de stage est justifiée car elle correspond à mettre en place les bases d'une collaboration internationale entre trois équipes de recherche, l'une dirigée par **Ira Assent** dans le département de science de calcul de l'université d'Arrhus au Danemark, la seconde dirigée par **Gert Cauwenberghs** de l'université de San Diego aux Etats Unis, et l'équipe IMAP du LISIC dirigée par **Christophe Renaud**. Cette collaboration naissante a été initiée au mois de mai 2017 lors d'un séjour de recherche dans l'université d'Arrhus où j'ai travaillé sur la thématique de l'apprentissage actif pour la détection d'anomalies avec Ira Assent. Puis, plus récemment, avec le professeur Gert Cauwenberghs de l'université de San Diego qui s'est proposé de collaborer sur la formulation incrémentale de l'apprentissage à une classe.

Programme et échéancier : la mise en œuvre se fera sur matlab à partir de la plateforme de calculs « calculco ». ce projet sera découpé en plusieurs étapes : une phase bibliographique permettant de faire le point sur la question, une phase de mise en œuvre et de validation sur des données synthétiques, une phase de test réalisée sur des flux vidéo réelles. La phase de rédaction conclura ce stage.