

Offre de stage

*Sujet : Détection des signaux faibles : application à
l'analyse des opinions*

Possibilité de poursuivre sur une thèse

Encadrement

Chloé Clavel , Anne Sabourin

Lieu et dates du stage

Telecom ParisTech, 46 rue Barrault, 75013 Paris

Date de début du stage : printemps 2016

Équipe(s) d'accueil de la thèse

département TSI, équipe STA-AAO <http://www.tsi.telecom-paristech.fr/sta/>.

Mots clés

analyse des opinions, détection de signaux faibles, machine learning, détection de Buzz

Sujet détaillé

L'analyse des opinions dans les réseaux sociaux est un domaine de recherche en plein essor, poussé d'une part, par les enjeux applicatifs du domaine (recommandation, e-reputation, gestion de la relation client), et d'autre part, par la multiplication des plate-formes d'expression des citoyens et des médias sur le web. Dans ce contexte, un des enjeux majeurs est de pouvoir détecter les signaux faibles afin de prédire l'arrivée d'un buzz sur les réseaux sociaux. L'objectif du stage sera de développer des méthodes de machine learning pour la détection de signaux faibles sur des données textuelles issues des réseaux sociaux. Des avancées récentes concernant l'apprentissage de la structure de dépendance des événements rares [1, 2] constituent un point de départ possible. En effet, les algorithmes développés dans ces travaux permettent d'identifier les groupes de features d'un jeu de données susceptibles de prendre des valeurs extrêmes simultanément: la question du pouvoir explicatif de ces

groupes concernant l'émergence potentielle d'un 'buzz' reste ouverte, c'est une des problématiques centrales de ce stage.

La Chaire Machine Learning for Big Data

Le traitement statistique des masses de données convoque à la fois mathématiques appliquées et informatique, à travers une discipline en plein essor : le Machine Learning ou apprentissage statistique.

Créée en septembre 2013 avec le soutien de la Fondation Télécom et financée à hauteur de près de 2 M€ par quatre entreprises partenaires : Criteo, PSA Peugeot Citroën, Safran et BNP Paribas, la Chaire Machine Learning for Big Data est portée par le mathématicien Stéphan Cléménçon, Enseignant-Chercheur, Professeur au sein du Département du Traitement du Signal et des Images à Télécom ParisTech.



Proposant cinq axes de recherche méthodologiques, enrichis par des applications industrielles concrètes, cette Chaire a pour objectif d'animer, en interaction avec ses partenaires, une activité de recherche de pointe en Machine Learning, ainsi que de proposer des programmes de formation.

La variété des données aujourd'hui disponibles (nombres, images, textes, signaux), leur grande dimension et leur volumétrie rendent souvent inopérantes les méthodes statistiques traditionnelles reposant sur le prétraitement humain et un long travail de modélisation. Le Machine Learning vise donc à élaborer et étudier des algorithmes, à vocation prédictive le plus souvent, permettant à des machines d'apprendre automatiquement à partir des données et à effectuer des tâches de façon performante.

Les avancées technologiques, l'omniprésence des capteurs (systèmes embarqués, objets connectés, Internet...) et l'explosion des réseaux sociaux s'accompagnent d'un véritable déluge de données, propulsant les sciences de l'information au centre du processus de valorisation des masses de données. En plus de la collecte et du stockage, l'enjeu est de pouvoir analyser ces données afin d'optimiser les décisions et mettre au point de nouvelles applications.

Au-delà du buzz médiatique dont il fait l'objet, le Big Data est donc un sujet stratégique majeur, au cœur d'enjeux économiques et sociétaux considérables. Son impact est désormais perçu dans presque tous les secteurs de l'activité humaine : de la recherche scientifique à la médecine en passant, entre autres, par la finance, le bâtiment, l'e-commerce, la défense ou les transports.

En savoir plus sur la Chaire, ses axes de recherche, ses activités, ses publications :

<http://machinelearningforbigdata.telecom-paristech.fr>

Profil du candidat

Etudiant titulaire d'un master 2 recherche

- Apprentissage statistique / traitement du Langage Naturel
- Bon niveau en programmation (Python ou R)
- Bon niveau d'anglais

Candidatures

À envoyer à :

chloe.clavel@telecom-paristech.fr, anne.sabourin@telecom-paristech.fr

- Curriculum Vitae
- Lettre de motivation personnalisée expliquant l'intérêt du candidat sur le sujet **(directement dans le corps du mail)**
- Relevés de notes des années précédentes
- Contact d'une personne de référence

Les candidatures incomplètes ne seront pas examinées.

Références

[1] Goix, N., Sabourin, A., Cléménçon, S. [Sparse Representation of Multivariate Extremes with Applications to Anomaly Ranking](#), AISTAT 2016

[2] Chiapino, M., Sabourin, A.(2016). [Feature clustering for extreme events analysis, with application to extreme stream-flow data](#) , ECML-PKDD 2016, workshop NFmcp2016

[3] Clavel, C.; Callejas, Z., Sentiment analysis: from opinion mining to human-agent interaction, Affective Computing, IEEE Transactions on, 7.1 (2016): 74-93.