

Offre de stage de Master (1 ou 2 ou élève ingénieur)

Titre du stage : Exploration de similitudes dans des corpus massifs temporels multimédia à base de graphes de connaissances
Mots-clés : machine-learning, deep-learning, explicabilité, biais algorithmique, mesure de similarité, images, réseaux sociaux
Laboratoire d'accueil : équipe EID du LIASD (EA 4383), Université Paris 8
Contacts Nédra Mellouli (n.mellouli@iut.univ-paris8.fr)
Durée : 4 à 6 mois, du 1/02 au 31/08/2022
Rémunération : 546 €/mois
Lieu du stage : Université Paris 8, IUT de Montreuil, 140, rue de la nouvelle France 93100 Montreuil

Le stage s'inscrit dans une recherche conduite au groupe de recherche EID du LIASD et en collaboration avec le Centre d'Expérimentation en méthodes numériques pour les recherches en SHS de Sorbonne Université

Contexte

La recherche d'information est un domaine de recherche en perpétuel mutation à cause de la transformation numérique. Le volume des données produit via les outils numériques connectés excède la capacité humaine de les traiter manuellement et un recours à l'intelligence artificielle (IA) devient incontournable. Toutefois, ces modèles intelligents issus d'IA doivent garantir au moins le passage à l'échelle, la compréhension des données hétérogènes et la catégorisation sémantique de ces données. Dans ce contexte, la recherche de similitudes au sein d'un large corpus multimédia où se mêlent images, vidéos et textes (structurés ou non) est une tâche très complexe.

Dans le cadre d'une recherche sur les controverses en ligne, nous nous intéressons à la circulation d'informations et de discours dans les réseaux sociaux numériques (Twitter, TikTok, Instagram par exemple) sous différentes formes (textes, images, vidéo). Plus spécifiquement, nous cherchons à capturer, caractériser, analyser et expliquer les déclinaisons iconographiques, textuelles et narratives des images à mesure qu'elles circulent dans les RSN.

Sujet et missions

Quand il s'agit d'une image, celle-ci peut subir différentes transformations faisant appel à de nombreux outils de traitement d'images. Ces traitements peuvent être simples et agissent par exemple sur les couleurs, ou encore sur la répartition spatiale des pixels (le recadrage, le découpage, etc.), ou bien plus complexes afférant à des outils : de génération d'images à partir de motifs d'une image de référence ; de substitution de l'image source ; de l'augmentation de l'image source par une caricature, un slogan, un

tag, une image codifiée, voire même une nouvelle image.

[?] Travail demandé

- cibler des événements clés sur les réseaux sociaux : constituer des corpus de référence ;
- élaborer un état de l'art sur les outils d'analyse de corpus massifs pluri-sémiotique : focus sur des corpus d'images ;
- étudier l'état des lieux bibliographique des mesures de similarité entre image, se focaliser sur les fonctions de similarité temporelle et contextuelle, et étudier en particulier le biais du choix de la mesure, ou de la fonction sur la qualité des résultats ;
- développer un modèle d'extraction semi-automatique d'objets caractéristiques d'un événement donné : les signatures iconographiques qui symbolisent l'événement par apprentissage. Le modèle se base sur une architecture de réseaux de neurones profonds à base de graphe de connaissances, et le plongement de graphes ;
 - augmenter le modèle par un algorithme de reconnaissance d'un événement à base de signatures ;
- implémenter et tester de nouvelles mesures de similarité qui soient à la fois robustes aux différences de cadrage, de résolution, de couleurs, de contraste, etc. et aux différentes évolutions, transformations, des objets caractéristiques au fil du temps de manifestation de l'événement étudié ;
- visualiser les schémas narratifs d'un objet caractéristique pour une collection d'événements de tests ;

Perspectives : Elargir ce travail dans le cadre d'une thèse pour le tracking des images fake : compréhension du processus de fabrication, génération et recommandation d'images de correction.

Prérequis

- Python, PyTorch, OpenCV, CUDA, gestion des données, machine-learning, deep-learning, visualisation

Pour candidater : Vous êtes étudiant.e dans un Master2 en informatique, science des données ou statistiques, ou étudiant.e dans une école d'ingénieur. Vous êtes enthousiaste pour la recherche, vous aimez comprendre en profondeur les problèmes et trouver des solutions élégantes. Vous avez une solide formation en informatique (Python pour l'environnement d'apprentissage automatique). Vous êtes intéressé.e par l'intelligence artificielle et, plus précisément, par l'apprentissage automatique, les techniques d'optimisation, l'analyse de données, etc. Vous avez un intérêt pour le domaine des SHS.

Candidature à adresser à n.mellouli@iut.univ-paris8.fr **ayant pour objet « Candidature Stage EID -Sim»** d'ici le **24 janvier 2022**, accompagnée d'un CV et d'une lettre de motivation, ainsi que vos notes de l'année universitaire précédente.

Les candidat·es retenu·es après examen des dossiers seront convoqué·es pour un entretien en visio-conférence avant le 28 janvier.

Bibliographie :

[1] Ilaria Tiddi, Stefan Schlobach. *Knowledge graphs as tools for explainable machine learning: A survey*, *Artificial Intelligence*, Volume 302, 2022,

[2] Jeremy Petch, Shuang Di, Walter Nelson. *Opening the black box: the promise and limitations of explainable machine learning in cardiology*, *Canadian Journal of Cardiology*, 2021,

[3] Satya M. Muddamsetty and Mohammad N. S. Jahromi and Thomas B. Moeslund. *SIDU: Similarity Difference and Uniqueness Method for Explainable AI*. In Computer Vision and Pattern Recognition 2020,

[4] A.Preece, D.Harborne, D.Braines, R.Tomsett, S.Chakraborty. *Stakeholders in Explainable AI*, 2018,

[5] Davide Castelvechi, Can We Open the Black Box of. AI?, NATURE, 2016,