Le laboratoire de Mathématiques et Informatique pour la Complexité et les Systèmes

Présente

L'AVIS DE SOUTENANCE

de **Monsieur Géraud Faye**
à l'école doctorale INTERFACES
CentraleSupélec, Université Paris Saclay, qui soutiendra publiquement ses travaux de thèse de doctorat intitulés :

« Misinformation Detection: Towards More Factual and Reliable Hybrid AI Approaches Over Time»

Sous la Direction de Madame Wassila Ourdane, Madame Céline Hudelot et l'encadrement de Monsieur Sylvain Gatepaille.

Le mercredi 8 avril à 14h

À l'école CentraleSupélec, en salle **E. 070 (théatre)** - Bâtiment Bouygues.

Membres du jury :

Vincent CLAVEAU, Cadre scientifique, AMIAD - Agence Ministérielle pour l'IA de Défense, Rapporteur,
Fabian SUCHANEK, Professeur des universités, Télécom Paris, Institut Polytechnique de Paris, Rapporteur,
Pascale SÉBILLOT, Professeure des universités, IRISA, Equipe de recherche Linkmedia, Examinatrice
Raphaël TRONCY, Professeur associé, EURECOM, Examinateur
Serena VILLATA, Directrice de recherche, Laboratoire I3S, CNRS, Examinatrice

Résumé :

La détection de la désinformation est une tâche critique dans le contexte géopolitique actuel. Le sujet est complexe et présente certaines spécificités que d'autres tâches de Traitement Automatique du Langage (TAL) n'exhibent pas, telles que le besoin d'évaluer des faits et de manipuler des connaissances. Dans ce travail, nous explorons la tâche de détection de la désinformation à travers différentes perspectives complémentaires, principalement centrées sur la façon dont les connaissances et les faits sont traités dans les systèmes de TAL :

(i) Dans une première partie, nous nous intéressons aux interactions entre textes et graphes pour rendre les méthodes d'encodage de texte plus orientées vers les faits et proposons TEG et TEGRA, améliorant les performances des classificateurs pour la détection de fausses informations en utilisant une représentation hybride texte-graphe.

\ (ii) Dans une deuxième partie et plus centré sur les faits, nous proposons FactNET, un nouveau cadre pour la vérification assistée des faits, permettant une analyse plus fine des affirmations complexes.

(iii) Dans une troisième partie, nous étudions la généralisation des classificateurs de désinformation lorsqu'ils sont confrontés à des connaissances futures, montrant un grand écart de performance par rapport à ce qui est mesuré dans des environnements contrôlés. Nous proposons LabDrift, une nouvelle métrique pour identifier les biais temporels dans les ensembles de données et deux approches pour améliorer la généralisation temporelle des modèles d'encodage.

(iv) Finalement, nous abordons la valorisation des modèles avec le développement d'une interface avec objectif l'éducation aux médias.

Abstract:

Misinformation detection is a critical task in current geopolitical context. The topic is complex and has some specificities that other Natural Language Processing (NLP) do not exhibit, such as the need for evaluating facts and manipulating knowledge. In this work, we explore the task of misinformation detection through different complementary lenses, mostly centered around how knowledge and facts are dealt with in NLP systems:

(i) In a first part, we are interested in interactions between texts and graphs to make text encoding methods more fact-oriented and propose TEG and TEGRA, improving performance of classifiers for fake news detection by using an hybrid text-graph representation.

(ii) In a second part and more centered on facts, we propose FactNET, a new framework for assisted fact-checking, allowing for a more fine-grained analysis of complex claims.

(iii) In a third part, we study the generalization of misinformation classifiers when confronted with future knowledge, showing a great gap in performance to what is measured in controlled environments. We propose LabDrift, a new metric to identify temporal biases in datasets and two approaches to improve temporal generalization of encoder models.

(iv) Finally, we worked on the valorization of models with the development of interface with the objective of developing media literacy.