Sujet thèse Agorantic 2021

Connaissance, analyse et contrôle de la propagation des fake news sur les réseaux sociaux

LIA et LBNC - Avignon Université

Laboratoires : Laboratoire Informatique d'Avignon et Laboratoire Biens, Normes, Contrats

Durée : 3 ans - **Début :** Septembre 2021

Salaire : 1500 euros/mois environ

Titre en anglais : Knowledge, analysis and control of the propagation of fake news on social networks.

Profil du candidat : Le demandeur doit posséder un Master en Informatique. Il doit maîtriser au moins un langage de programmation objet courant (Java, C++...) et un langage de script (Python, Perl...). En outre, une expérience dans une des thématiques liées au sujet (traitement automatique de langage, fouille de données, apprentissage automatique, réseaux complexes...) serait appréciée. Des bases solides en mathématiques sont également attendues. Il devra enfin montrer un intérêt sérieux à la pluridisplinarité, et s'intéresser en particulier à l'aspect juridique et sociétal du traitement de la fake news, qui est au cœur de ce sujet de thèse.

Les candidatures doivent être adressées à :
- Richard Dufour (richard.dufour@univ-avignon.fr) - LIA, Université d'Avignon
- Emmanuel Nether (emmanuel.netter@univ-avignon.fr) - LBNC, Université d'Avignon
- Rachid Elazouzi (rachid.elazouzi@univ-avignon.fr) - LIA, Université d'Avignon

et doivent inclure :
- un CV détaillé (formation et expériences en recherche),
- une lettre de motivation spécifiant de façon détaillée la motivation du candidat quant au sujet interdisciplinaire proposé,
- les notes de Licence et de Master dans tous les modules suivis,
- au moins une référence qui peut être contactée pour recommandation.

Mots clés : Traitement du langage, Réseaux complexes, Cadre juridique, Droit des réseaux sociaux, Propagation de l'information, Corpus.

Contexte général

La diffusion de fausses nouvelles (fake news) n'est pas un phénomène nouveau, mais est présente tout au long de l'Histoire de l'humanité : les historiens en retrouvent notamment des traces depuis l'Antiquité. Ces fake news, qui peuvent émaner d'une ou plusieurs entités à des niveaux institutionnels différents (personne isolée, média, gouvernement...), visent à désinformer, souvent dans un objectif précis (déstabilisation d'un gouvernement, recherche d'avantages financiers, trucage d'élections...). Ce phénomène a pris une ampleur encore inédite ces dernières années, en particulier avec l'avènement des réseaux sociaux, qui permettent d'échanger des

informations très rapidement à une très grande masse de personnes sans vérification et régulation de ces dites informations.

Même pour des utilisateurs avertis des réseaux sociaux, les fake news sont très difficiles à identifier puisque, par essence, elles ont été créées pour paraître le plus réaliste possible. Un des objectifs serait de pouvoir les détecter automatiquement, et ce le plus tôt, afin d'éviter leur propagation ou de les rectifier si nécessaire. Des événements d'ampleur mondiale, tels que le fake news challenge (http://www.fakenewschallenge.org) en 2017 [hanselowski2018], ou encore plus récemment le Fake News Detection Challenge KDD en 2020 (https://www.kaggle.com/c/fakenewskdd2020), ont mobilisé la communauté scientifique pour faire avancer la recherche dans la détection automatique des fake news. L'ampleur des travaux pour sa détection automatique, et leur rapide évolution, est telle que de nombreuses publications scientifiques se sont attelées à les résumer ces dernières années [oshikawa2018, sharma2019, bondielli2019, zhou2020]. De nombreux domaines de recherche sont alors impliqués, allant assez naturellement du traitement automatique du langage, avec extraction de caractéristiques textuelles [bondielli2019], en passant par la théorie des graphes et les réseaux complexes [liu2018], ou encore l'analyse d'images et leur manipulation [huh2018]. Des approches combinant différentes sources d'information sont également étudiées [wang2018]. Les caractéristiques extraites permettant de détecter automatiquement les fakes news peuvent lors être multiples, alors du contenu linguistique (représentation des mots, mots-clés liés aux émotions, descripteurs du discours et de la syntaxe...), d'indices visuels (cohérence de l'image, mesures de similarité...), de réseaux d'information (graphes d'interaction, liens d'amitié entre utilisateurs, timeline...).

Des approches de natures différents sont également souvent évoquées, comme par exemple dans [zhou2020], où les auteurs ont choisi de catégoriser les méthodes selon qu'elles soient à base de connaissances (vérification de la véracité par rapport au contenu), qu'elles s'appuient sur la propagation de la nouvelle au sein d'un réseau, qu'elles la vérifient au sein de sources extérieures (fact checking), ou enfin qu'elles analysent le style de rédaction de la nouvelle (i.e. de la façon dont elle est écrite).

Malgré les avancées pour leur détection automatique, et les efforts fournis par le milieu de la recherche, des verrous scientifiques persistent. Pour les approches s'appuyant sur le contenu textuel en traitement du langage, il est très difficile de proposer des modèles permettant de généraliser à de nouvelles fakes news, dont le vocabulaire peut évoluer tout comme les indices pouvant être identifiés (ceux-ci peuvent varier selon la thématique abordée par exemple). Cela est également rendu difficile par la nature même de la fake news, qui, par essence, tend à proposer des "fausses" nouvelles extrêmement ressemblantes à des vraies. Permettre de vérifier les nouvelles pose aussi le problème d'accès à des ressources extérieures, et de pouvoir les analyser et "comprendre" automatiquement. Enfin, les approches n'utilisant pas le contenu textuel, comme par exemple les réseaux de propagation des actualités, apparaissent comme plus robustes, mais peuvent souffrir de la taille des réseaux à traiter (par exemple, les masses de données gigantesques échangées sur les réseaux sociaux), mais également du fait que ces fake news ne seraient détectées qu'une fois celles-ci propagées.

Les réseaux sociaux sont des environnements spécifiques mis en tension par des enjeux à la fois internes propres à la constitution du champ et externes propres à la société dont il est un relais d’information. Les enjeux internes concernent notamment les règles d’accès et la régulation qui sont mises en place par les dirigeants des réseaux sociaux. Ces règles sont souvent modifiées par le cadre dans lequel elles s’inscrivent et qui affectent en retour leurs activités. Facebook et

Twitter sont les agents sociaux à la pointe de cette autonomisation. D'autres acteurs comme les médias (entreprises de presse, etc.), intègrent les dispositifs d'information et de communication du web dans leurs champs respectifs, en vertu de logiques qui leur sont propres. Dans ce cadre, les réseaux sociaux sont considérés comme dispositif technique à partir duquel se déploient dans la durée des acteurs qui partagent un objectif commun. L'évolution récentes des réseaux sociaux a permis d'émerger des nouveaux acteurs extrêmement actifs et coordonnés pour propager les fakes news. Les travaux sur les fake news ont souvent ignoré l'aspect stratégique de ces acteurs et ont considéré que la propagation des fake news pourrait être modélisée comme un phénomène du Buzz. Il convient donc de prendre en compte ces transformations pour construire des solutions efficaces pour contenir la viralité des fake news. L'identification et l'unification de ces acteurs au travers de leurs identités explicites ou implicites et des réseaux qu'ils tissent est un problème difficile que l'on doit envisager conjointement avec l'identification des sources.

Un dernier verrou concerne l'aspect juridique lié au contrôle des fake news. En effet, les efforts faits ces dernières années concernent principalement leur détection et leur explication. Cependant, d'un point de vue juridique, il apparaît difficile de supprimer a priori un contenu, ce qui touche directement au problème de la liberté d'expression. Il est plus facile en effet d'attendre d'une plateforme de réseau social qu'elle supprime les contenus "haineux" les plus violents, car leur contrariété au droit français est manifeste. Pour qualifier un contenu de "fake news", en revanche, il faut caractériser à la fois leur fausseté, ce qui peut être délicat dans des débats complexes, mais également une intention de manipuler le lecteur ("fake news" étant volontiers traduit par les juristes comme *informations trompeuses* plutôt que simplement *fausses*). Le projet de règlement européen Digital Services Act fait pourtant obligation aux plateformes d'identifier et de limiter tous les « risques systémiques » susceptibles de survenir sur leurs services. Or, la diffusion d'informations trompeuses, par exemple à des fins de manipulation électorale, entre nettement dans cette catégorie. Il convient donc de prendre en compte ces aspects juridiques, sociaux et sociétaux pour construire des modèles et des solutions qui leur permettent de sortir des milieux scientifiques dans lesquels ces approches sont issues.

Objectifs scientifiques

La lutte contre les fake news et pour la qualité de l'information passe par des mesures propres à limiter l'influence des médias de masse et des médias sociaux. L'objectif principal de la thèse est d'étudier les différentes possibilités d'actions contre la propagation nuisible des fausses informations. Les approches existantes ne permettent pas une bonne intégration entre l'analyse des propagation et l'évaluation des sources de l'information. Ce problème est causé par la limitation de l'analyse des cascades d'informations à l'intérieur d'un seul réseau social. Dans cette thèse, nous souhaitons élaborer un modèle des cascades de désinformation sur les réseaux sociaux en utilisant les caractéristiques qui sont propres à un réseau social particulier. Les verrous scientifiques sont très divers et ils relèvent différents domaines scientifiques :

- Caractériser et identifier une fake news d'un point de vue multimodal semble essentiel pour arriver à un système suffisamment performant pour être applicable massivement. Cela passe bien entendu par l'application de méthodes issues du traitement automatique du langage, en particulier au moyen d'architectures par apprentissage profond à l'état de l'art (BERT, Elmo…), mais également l'intégration d'informations extra-linguistiques (profil de l'utilisateur, réseaux d'interaction, précédents relais d'informations "suspectes"...) à intégrer le plus tôt possible dans ces modèles.

- En prolongement du premier objectif, il s'agira non pas de travailler sur l'individu lui-même mais d'être capable d'identifier les personnes ou groupes susceptibles d'être affectés par une information. En effet, il semble que la cible des fake news dépasse clairement le cercle d'amitié liant déjà les individus : il faut convaincre des personnes indécises, ou au moins leur faire douter sur certains sujets. L'analyse des contenus et des acteurs joue un rôle important pour répondre à cette question, où il faudra être capable d'estimer, autant que possible, l'impact des fake news sur les utilisateurs.
- Identifier et extraire les paramètres caractérisant ou contribuant à la propagation des fakes news : la façon dont l'événement est évoqué, le volume de l'information et les canaux empruntés. L'étude de la structure et la propagation de l'information est à l'interface de deux disciplines : théorie des graphes et l'épidémiologie.
- Offrir une modélisation réaliste de la diffusion de l'information en développant des modèles mathématiques en tenant en compte de plusieurs paramètres : Relais Media (*trad on online*), nature de l'information, l'influenceur et les réseaux sociaux utilisés.
- Proposer des mesures à l'échelle des réseaux sociaux pour limiter la propagation des fake news sans passer par un blocage strict des contenus. Une étude juridique sera aussi proposée pour comprendre l'applicabilité de nos solutions par rapport à l'application du droit de la concurrence aux *fake news*.

Références bibliographiques :

[Bondielli19] Bondielli, A., & Marcelloni, F. (2019). A survey on fake news and rumour detection techniques. Information Sciences, 497, 38-55.

[Hanselowski18] Hanselowski, A., PVS, A., Schiller, B., Caspelherr, F., Chaudhuri, D., Meyer, C. M., & Gurevych, I. (2018). A retrospective analysis of the fake news challenge stance detection task. arXiv preprint arXiv:1806.05180.

[Huh18] Huh, M., Liu, A., Owens, A., & Efros, A. A. (2018). Fighting fake news: Image splice detection via learned self-consistency. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 101-117).

[Liu18] Liu, Y., & Wu, Y. F. (2018, April). Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 32, No. 1).

[Oshikawa18] Oshikawa, R., Qian, J., & Wang, W. Y. (2018). A survey on natural language processing for fake news detection. arXiv preprint arXiv:1811.00770.

[Sharma19] Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). Combating fake news: A survey on identification and mitigation techniques. ACM Transactions on Intelligent Systems and Technology (TIST), 10(3), 1-42.

[Wang18] Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., ... & Gao, J. (2018, July). Eann: Event adversarial neural networks for multi-modal fake news detection. In Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining (pp. 849-857).

[Zhou20] Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Surveys (CSUR), 53(5), 1-40.