

Intégration d'une méthode d'explicabilité pour l'analyse d'opinions sur les médias sociaux

Mots clés : Explicabilité, XAI, analyse des médias sociaux, fouille d'opinions

Encadrant : Maria Malek

Laboratoire : ETIS UMR 8051

Équipe : MIDI

Contexte

Les racines de l'IA remontent à plusieurs décennies, les méthodes d'apprentissage automatique atteignent aujourd'hui des niveaux de performance élevées permettant de résoudre des tâches de plus en plus complexes. Il existe aujourd'hui un besoin émergent de comprendre comment ces décisions sont rendues par des méthodes d'IA lorsque les décisions de tels systèmes affectent la vie des humains. Les paradigmes sous-jacents à ce problème révèlent du domaine dit de l'IA explicable (XAI) reconnu comme une caractéristique cruciale pour le déploiement pratique des modèles d'IA [1]. L'explicabilité a pour objectif de faciliter la compréhension de divers aspects d'un modèle conduisant à des informations qui peuvent être utilisées par plusieurs acteurs. L'explicabilité peut être considérée comme une caractéristique active d'un modèle, désignant toute action prise par un modèle dans le but de clarifier ou de détailler ses fonctions internes [1,2].

D'un autre côté, l'étude des réseaux complexes nommés aussi graphes de terrain est un domaine de recherche scientifique jeune et actif largement inspiré des découvertes empiriques sur les réseaux réels tels que les réseaux informatiques, les réseaux biologiques, les réseaux technologiques, les réseaux cérébraux, les réseaux climatiques et les réseaux sociaux [3].

Nous explorons dans nos travaux actuels [5] portant sur l'analyse des médias sociaux, la combinaison de méthodes classiques d'exploration d'opinion avec l'analyse des réseaux sociaux et son impact sur la formation et la propagation d'opinion afin de construire un modèle d'opinion cohérent.

Afin d'étudier l'impact des utilisateurs influents (nœuds influents), nous intégrons dans un premier temps plusieurs facteurs d'influence extraits du réseau dans le processus d'exploration d'opinions. Ces facteurs sont généralement calculés en utilisant différentes mesures de centralité comme le degré, la proximité, l'intermédiarité, la centralité PageRank, etc.

Nous définissons et étudions ensuite la notion de la stabilité d'opinion au sein des réseaux égocentriques autour des influenceurs et au sein des communautés détectées [4], notre objectif étant de détecter la modification d'opinion pour les deux types de sous-réseaux. Nous analysons les communautés obtenues afin de comprendre les opinions émergentes à partir de ces communautés non seulement en fonction des profils utilisateurs mais aussi en fonction d'éléments topologiques. Nous souhaitons également proposer des indicateurs concernant la stabilité des opinions et d'autres liés à leurs changements.

Objectif

Le but de stage est de proposer et d'intégrer une méthode d'explicabilité dans les algorithmes d'analyse d'opinions afin de produire des explications émergentes qui combinent des informations nodales (comme le profil d'utilisateur) et topologiques extraites de la structure du graphe de propagation des opinions.

En intégrant une méthode d'explicabilité adéquate, nous souhaitons rendre plus compréhensible également les résultats concernant la polarité de l'opinion trouvée au niveau des utilisateurs et au niveau des groupes. De même, le modèle doit être capable d'expliquer les changements d'opinion détectés en lien avec les informations extraites du réseau de propagation et les séquences d'actions entreprises (par exemple : tweets, retweets, réponses) menant à ce changement.

Dans la plupart des méthodes d'apprentissage automatique les entités traitées dans une méthode d'explicabilité sont liées aux caractéristiques (attributs), aux données, aux prototypes extraits ainsi qu'à des connaissances simples extraites ayant souvent la forme de règles ou d'arbre décision. Lorsqu'il s'agit des réseaux complexes, des nouvelles entités doivent être prises en compte comme la nature des liens dans le graphe ainsi que les informations topologiques.

Références

[1] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador Garcia, Sergio Gil-Lopez, Daniel Molina, Richard Benjamins, Raja Chatila, Francisco Herrera, Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, Information Fusion, Volume 58, 2020, Pages 82-115, ISSN 1566-2535, <https://doi.org/10.1016/j.inffus.2019.12.012>

[2] Andrea Morichetta, Pedro Casas, and Marco Mellia. 2019. EXPLAIN-IT: Towards Explainable AI for Unsupervised Network Traffic Analysis. In Proceedings of the 3rd ACM CoNEXT Workshop on Big DATA, Machine Learning and Artificial Intelligence for Data Communication Networks (Big-DAMA '19). Association for Computing Machinery, New York, NY, USA, 22–28. DOI:<https://doi.org/10.1145/3359992.3366639>

[3] J. Newman, The structure and function of complex networks, SIAM Reviews, 45(2): 167-256, 2003

[4] S. Fortunato, Community detection in graphs, Physics Reports 486, 75-174 (2010), <http://dx.doi.org/10.1016/j.physrep.2009.11.002>

[5] Kossi Folly, Maria Malek, Dimitris Kotzinos. Social networks analysis for opinion model extraction. *Networks 2021: first combined meeting of the International Network for Social Network Analysis (Sunbelt XLI), and the Network Science Society (NetSci 2021).*, Jul 2021, Indiana, United States.

Compétences requises : connaissances en apprentissage automatique et compétences en Python.

Candidature : lettre de motivation, CV et relevés de notes M1/M2

Contact : maria.malek@cyu.fr