

Titre: Inférence de Réseaux à Partir des Données Hétérogènes avec Graph Neural Networks

Mots-clés

Réseaux Complexes, Inférence de Réseaux, Graph Neural Networks, Dynamique de Transmission, Modélisation Spatio-temporelle, Influenza Aviaire

Profil recherché

Master ou diplôme d'ingénieur en informatique ou en mathématiques appliquées liées à l'apprentissage automatique/statistique. Le candidat doit montrer un certain intérêt pour les aspects théoriques de l'apprentissage automatique/des réseaux de neurones sur graphes, ainsi que des compétences en programmation Python. En outre, il doit être capable de documenter le travail réalisé de manière structurée et claire en anglais.

Durée

5 à 6 mois, à partir de Février jusqu'en Août 2025

Rémunération

Selon le barème officiel en vigueur, environ 650 €/mois

Lieu

INRAE, UMR TETIS
Maison de la Télédétection, 500 rue Jean François Breton, Montpellier

Encadrants

Nejat Arinik: nejat.arinik@univ-artois.fr
Roberto Interdonato: roberto.interdonato@cirad.fr
Maguelonne Teisseire: maguelonne.teisseire@inrae.fr

Pour candidater

Merci d'envoyer un mail à nejat.arinik@univ-artois.fr et roberto.interdonato@cirad.fr avec sujet "CANDIDATURE STAGE MOOD 2025".

- lettre de motivation expliquant vos qualifications, expériences et motivation pour ce sujet (1-2 pages)
- curriculum vitae (1-2 pages)
- relevé de notes de 1ère année de master et les notes de 2ème année de master disponibles ou équivalent pour les écoles d'ingénieurs
- un lien vers des dépôts de projets personnels (par exemple GitHub ou GitLab)
- toute autre information que vous estimerez utile

Contexte

Ce stage de master est proposé dans le cadre du projet MOOD (The MOonitoring Outbreaks for Disease surveillance in a data science context). Un objectif particulier de ce projet est d'étudier la dynamique de transmission de l'influenza aviaire (IA), qui est une maladie animale très contagieuse [1]. Bien que quelques cas d'infection humaine aient eu lieu dans le passé, elle infecte principalement les espèces d'oiseaux sauvages et domestiques. La transmission entre les oiseaux peut être directe, par contact étroit, ou indirecte, par l'intermédiaire de matériaux contaminés tels que l'eau et les aliments. Les oiseaux sauvages migrateurs, en particulier, jouent un rôle clé dans cette transmission et permettent aux virus de se propager sur de longues distances. L'émergence et la propagation de l'IA ont de graves conséquences pour la santé animale et un impact socio-économique considérable pour les producteurs de volailles du monde entier. En raison de sa nature hautement contagieuse, il est essentiel de surveiller l'évolution et la propagation de cette maladie.

En pratique, il est difficile de découvrir les véritables voies de transmission des virus de l'IA, c'est-à-dire de savoir d'où vient une épidémie qui apparaît à un endroit donné [2]. Néanmoins, l'acquisition de ces trajectoires de transmission peut permettre de découvrir des schémas de transmission et de prévoir à temps les épidémies suivantes. Dans ce travail, nous nous concentrons sur le problème de l'inférence de la manière dont la maladie IA se propage en l'absence de connaissances sur les voies de transmission réelles de la maladie. Nous l'abordons comme un problème d'inférence de réseau, en nous appuyant sur des données hétérogènes provenant d'un ensemble de données spatio-temporelles sur l'IA, accompagnées des séquences génétiques, et de données sur l'abondance d'oiseaux.

Méthodes

Il existe principalement deux groupes de travaux dans la littérature sur l'inférence de réseaux avec des données spatio-temporelles. Le premier groupe de travaux s'appuie essentiellement sur les méthodes d'inférence de réseaux à partir des traces temporelles d'événements infectieux, mais adaptées pour inclure la connectivité spatiale des entités spatiales (par exemple, les régions voisines, comme le fait [3]). Pour mieux saisir la dépendance spatiale avec une structure de graphe, les méthodes basées sur Graph Neural Networks (GNN, réseaux de neurones sur graphes en français) sont des approches prometteuses [4]. Ces méthodes peuvent être manipulées pour effectuer la tâche d'inférence

du réseau, souvent en décodant les vecteurs de caractéristiques des nœuds latents en arêtes prédites. Par exemple, [6] s'appuie sur les réseaux de neurones récurrents à barrière (GRU) pour modéliser les informations de l'épidémie Covid-19 dans les séries temporelles et sur Graph Convolutional Networks (GCN, les réseaux convolutifs pour graphe en français) pour obtenir la relation topologique entre la ville et ses voisins. Le deuxième groupe, appelé apprentissage de graphe basé sur un modèle [8], utilise des modèles de données liés aux données observées afin de formuler la tâche d'inférence de réseau comme un problème d'optimisation (convexe) [9]. Par exemple, [5] adopte un modèle de diffusion spatio-temporel et un framework d'analyse de survie pour estimer la structure du réseau des cas de paludisme.

Missions

Dans ce travail, nous visons essentiellement à inférer une structure de réseau statique en deux étapes en nous appuyant à la fois sur le GNN et sur des travaux fondés sur des modèles, mais adaptés à notre contexte de données épidémiologiques hétérogènes. Premièrement, il est essentiel d'appliquer les méthodes de l'état de l'art, qui concernent uniquement les attributs spatio-temporels de nos données épidémiologiques afin d'identifier leurs forces et leurs faiblesses. Même avec cette restriction de données, nous devons adapter les travaux existants à notre contexte en tenant compte du fait que la dynamique de l'IA est différente de celle des maladies humaines comme COVID-19 (par exemple, les interactions entre les oiseaux domestiques et sauvages, les trajectoires des oiseaux n'étant pas comme un réseau routier classique). Nous évaluerons ces méthodes sur la base d'ensembles de données synthétiques réelles et fondées sur des simulations. Deuxièmement, nous nous concentrons uniquement sur les meilleures méthodes d'inférence trouvées au cours de la première étape et nous les étendons pour inclure les informations de séquençage du génome et les données d'occurrence des oiseaux. En fonction des progrès réalisés au cours du stage, nous pouvons également étendre ce travail en inférant une structure de réseau dynamique, c'est-à-dire une topologie de réseau variant dans le temps, comme dans [7, 10], et une structure de réseau multicouche, comme dans [11], lorsque l'on prend en compte différentes souches d'IA (par exemple H5N1, H5N8) et différents types d'oiseaux (par exemple domestiques, sauvages).

Bibliographie

- [1] I. Brown, P. Mulatti, K. Smietanka, C. Staubach, P. Willeberg, C. Adlhoch, D. Candiani, C. Fabris, G. Zancanaro, J. Morgado, F. Verdonck (2017). [Avian influenza overview october 2016-august 2017](#), EFSA Journal 15, e05018.
- [2] D. H. Lee, M. K. Torchetti, J. Hicks, M. L. Killian, J. Bahl, M. Pantin-Jackwood, D. E. Swayne (2018). [Transmission Dynamics of Highly Pathogenic Avian Influenza Virus A\(H5Nx\) Clade 2.3.4.4, North America, 2014-2015](#). Emerg Infect Dis. (10):1840-1848.
- [3] B. Rozemberczki, P. Scherer, O. Kiss, R. Sarkar, T. Ferenci (2021). [Chickenpox Cases in Hungary: a Benchmark Dataset for Spatiotemporal Signal Processing with Graph Neural Networks](#). ArXiv, abs/2102.08100.

- [4] B. Rozemberczki, P. Scherer, Y. He, G. Panagopoulos, A. Riedel, M. Astefanoaei, O. Kiss, F. Beres, G. López, N. Collignon, R. Sarkar (2021). [PyTorch Geometric Temporal: Spatiotemporal Signal Processing with Neural Machine Learning Models](#). In Proceedings of the 30th ACM International Conf. on Information & Knowledge Management, 4564-4573.
- [5] I. Routledge, H. J. T. Unwin, S. Bhatt (2021). [Inference of malaria reproduction numbers in three elimination settings by combining temporal data and distance metrics](#). Sci Rep. 11(1):14495.
- [6] S. Yu, F. Xia, S. Li, M. Hou, Q. Z. Sheng (2023). [Spatio-temporal Graph Learning for Epidemic Prediction](#). ACM Trans. Intell. Syst. Technol. 14, 2.
- [7] J. Chen, X. Wang, X. Xu. [GC-LSTM: graph convolution embedded LSTM for dynamic network link prediction](#). Appl Intell 52, 7513-7528 (2022).
- [8] M. Wasserman, G. Mateos (2022). [pyGSL: A Graph Structure Learning Toolkit](#), arXiv:2211.03583.
- [9] G. Mateos, S. Segarra, A. G. Marques, A. Ribeiro (2018). "[Connecting the Dots: Identifying Network Structure via Graph Signal Processing](#)." IEEE Signal Processing Magazine, 36(3), 16-43.
- [10] L. Yu, L. Sun, B. Du, W. Lv. (2024). [Towards better dynamic graph learning: new architecture and unified library](#). In Proceedings of the 37th International Conference on Neural Information Processing Systems, 67686-67700.
- [11] Y. Zhao, X. Yao, H. Huang (2024). [A continuous-time diffusion model for inferring multi-layer diffusion networks](#). Appl Intell 54, 8200-8223.