

Sujet de TER : Deep learning pour l'extraction d'informations dans des bases de données géographiques

Encadrement : Gabriel Frey (g.frey@unistra.fr)

Aurélien Leborgne (aurelie.leborgne@unistra.fr)

Mots-clés : Réseaux de neurones, graphes, données spatio-temporelles.

Problématique : De nos jours, les données à analyser sont de plus en plus nombreuses. Certaines de ces données se modélisent naturellement sous la forme de graphes. En particulier, cela permet de facilement mettre en valeur l'existence de liens ou de sous-structures communes. Pour extraire de l'information de ces graphes, il est indispensable de pouvoir analyser efficacement ces ensembles de données. Pour cela, il est possible d'utiliser les réseaux de neurones. Dans de nombreux domaines où les données sont représentées sous forme de graphes, l'apprentissage d'une métrique de similarité entre les graphes est considéré comme un problème clé, ce qui peut faciliter davantage diverses tâches d'apprentissage, telles que la classification, le regroupement et la recherche de similitude. Récemment, il y a eu un intérêt croissant pour l'utilisation de réseaux de neurones pour estimer la similarité entre graphes, où l'idée clé est d'apprendre un modèle d'apprentissage en profondeur qui mappe les graphes d'entrée à un espace cible (plongement de graphe) de telle sorte que la distance dans l'espace cible se rapproche de la distance structurelle dans l'espace d'entrée. Pour ce faire, les Graph Neural Networks (GNN) peuvent être utilisés, comme le montre la Figure 1.

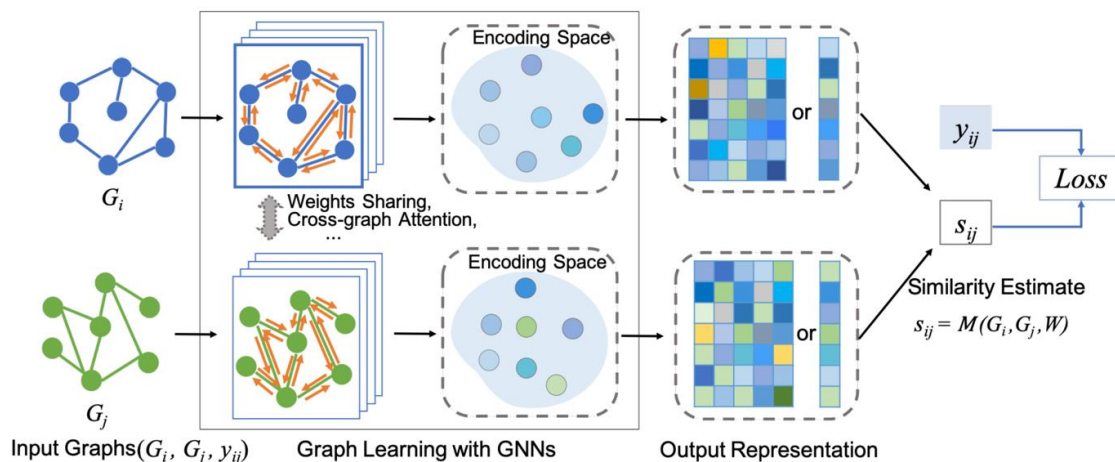


Figure 1: Illustration de l'apprentissage de la similarité de graphe basé sur GNN

Contexte : Les données étudiées seront l'évolution temporelle de zones géographiques. Pour ce faire, les données à votre disposition seront le Registre Parcellaire Graphique de la France entre 2015 et 2019 [1]. Ces données peuvent être représentées sous la forme de graphe.

Objectifs : Les objectifs scientifiques sont l'utilisation et la comparaison de méthodes GNN sur les graphes obtenus à partir de données géographiques afin de classer les graphes en fonction des caractéristiques des parcelles agricoles. En d'autres termes, il s'agira de regrouper les occupations de terrain qui se ressemblent et d'en extraire des caractéristiques. Pour cela, il sera nécessaire :

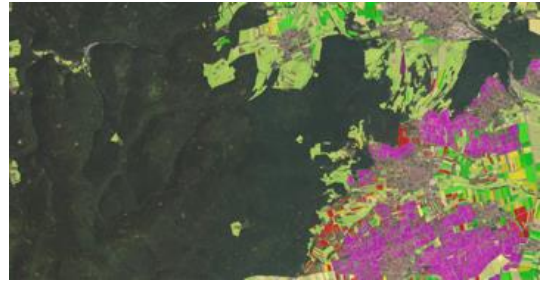


Figure 2 : Extrait du RPG dans les Vosges.

- d'effectuer un état de l'art des méthodes existantes [2] [3],
- de se familiariser avec les données géographiques et les traduire en graphes,
- de transformer les graphes en vecteurs,
- d'appliquer des méthodes de réseaux de neurones sur ces vecteurs,
- d'analyser ces résultats.

Un certain nombre d'implémentations sont disponibles et pourront être adaptées (par exemple [4]). En fonction des résultats obtenus, nous proposerons des **améliorations** afin que, dans l'idéal, nous puissions **publier un article** sur ce sujet dans lequel vous apparaîtrez bien évidemment comme auteur.

N'hésitez pas à nous contacter pour toutes questions (aurelie.leborgne@unistra.fr ou g.frey@unistra.fr) et à nous proposer un entretien afin de discuter ensemble de vos attentes et de nos attentes.

Courte Bibliographie :

[1] <https://geoservices.ign.fr/documentation/diffusion/telechargement-donnees-libres.html#rpg>

[2] MA, Guixiang, AHMED, Nesreen K., WILLKE, Theodore L., et al. Deep graph similarity learning: A survey. Data Mining and Knowledge Discovery, 2021, p. 1-38.

[3] WU, Zonghan, PAN, Shirui, CHEN, Fengwen et al, A Comprehensive Survey on Graph Neural Networks, CoRR, 2019

[4] <https://docs.dgl.ai/tutorials/blitz/index.html>