

Modélisation de parcours patients : graphes temporels pour la supervision médicale

Hugo Le Baher^{*,**,***}, Jérôme Azé^{*}, Sandra Bringay^{*,****}, Pascal Poncelet^{*}, Nancy Rodriguez^{*}, Caroline Dunoyer^{**,#}

- * LIRMM, UMR 5506, Université de Montpellier, CNRS, Montpellier, France
prenom.nom@lirmm.fr
- ** Département d'Information Médicale, CHU Montpellier, Montpellier, France
- *** 5 DEGRÉS, Paris, France
<https://www.5degres.com>
- **** AMIS, Université Paul-Valéry, Montpellier, France
prenom.nom@univ-montp3.fr
- # IDESP, UMR UA11, INSERM - Université de Montpellier, Montpellier, France
prenom.nom@umontpellier.fr

Résumé. L'usage de méthodes d'apprentissage automatique se démocratise pour anticiper les risques critiques chez les patients sous surveillance et diminuer la charge des soignants. Dans cet article, nous proposons une modélisation originale qui bénéficie des développements récents en convolution de graphes : un parcours patient est vu comme un graphe, où chaque nœud est un évènement et où les proximités temporelles sont représentées par des arcs pondérés. Cette modélisation a été évaluée pour prédire le décès à 24 heures sur un jeu de données réelles puis comparée avec succès avec les résultats de l'état de l'art.

1 Introduction

La supervision de l'évolution de la santé des patients dans les services d'urgence médicale est un problème difficile. Jung et al. (2016) montrent que la mise en place d'outils d'aide à la décision, qui sont en mesure d'anticiper les risques critiques chez les patients sous surveillance, améliorerait la capacité des services à réagir le plus vite possible aux épisodes inattendus et ainsi prévenir les décès de leurs patients.

Différentes approches ont été proposées pour prédire l'état d'un patient à partir de son historique. Généralement, toutes les données sont transformées sous la forme d'un vecteur dans le but d'utiliser des techniques traditionnelles d'apprentissage comme le partitionnement (Poongodi et al., 2021), la méthode des k plus proches voisins, la régression logistique, ou des modèles d'apprentissage profonds (El-Rashidy et al., 2020). Sous l'hypothèse que les variations dans le temps des descripteurs du patient sont essentielles pour prédire son état futur, d'autres techniques considèrent un parcours patient comme une séquence, constituée d'une succession de vecteurs. Cette nouvelle représentation peut être utilisée dans des modèles récurrents comme LSTM (Ashfaq et al., 2019; Li et al., 2020). La mise en place de ces techniques

nécessite de découper l'historique du patient selon la temporalité. La recherche du découpage qui représente le mieux toutes les données capturées dans le temps est un processus coûteux, qui influence directement la capacité de prédiction du modèle. Par exemple, si on considère la granularité de l'heure, nous perdons l'évolution du rythme cardiaque seconde par seconde. Au contraire, si nous considérons un découpage avec une granularité à la seconde, la représentation produite est trop volumineuse pour être prise en compte par la plupart des modèles et très peu d'évolutions seront observables d'un vecteur à l'autre. Un autre problème sous-jacent de cette représentation est que les différentes séquences des patients ont des longueurs variables ce qui peut nécessiter des adaptations dans le modèle d'apprentissage.

Récemment, les réseaux neuronaux de graphes (GNN - *Graph Neural Network*) ont montré qu'ils étaient très adaptés pour faire de la prédiction (Zhou et al., 2020). En vue des objectifs mentionnés précédemment et des limites identifiées des approches de la littérature, nous proposons dans cet article une nouvelle modélisation des dossiers patients sous la forme de graphes.

Les contributions de l'article sont les suivantes :

- Nous proposons une nouvelle modélisation originale des données temporelles et hétérogènes sous la forme d'un graphe où chaque nœud correspond à une mesure et où les proximités temporelles sont représentées via des arcs pondérés.
- Nous utilisons cette représentation en entrée d'un réseau à convolutions de graphes (GCN - *Graph Convolutional Network*) que nous évaluons sur le jeu de données réelles MIMIC-III décrit par Harutyunyan et al. (2019), pour prédire, à partir des données recueillies durant leur parcours, le décès des patients à 24h. Nous montrons que notre approche offre des résultats comparables à ceux de l'état de l'art.

Le reste de l'article est organisé de la manière suivante. La section 2 décrit la modélisation proposée pour intégrer directement la dimension temporelle. La section 3 décrit les expérimentations menées. Enfin, nous concluons en présentant nos travaux futurs.

2 Modélisation

Dans cet article, nous proposons une modélisation des données issues des dossiers patients sous forme de graphes qui a pour objectif de ne pas matérialiser de découpage temporel.

Formellement, le parcours d'un patient est représenté par un graphe $G = (E, P, X)$ où $E = \{e_1, e_2, \dots, e_n\}$ représente les événements, $P = \{(e_i, e_j, p)\}, 1 \leq i \leq n, 1 \leq j \leq n, p \in \mathbb{R}^+$ représente l'ensemble des arcs pondérés avec la proximité temporelle p entre deux événements e_i et e_j et $X \in \mathbb{R}^{n \times m}$ représente les caractéristiques des événements avec m le nombre de modalités distinctes mesurées.

Notre objectif est de prédire le résultat médical d'intérêt, ici le décès à 24h. Il s'agit d'un problème de classification binaire consistant donc à trouver un modèle f tel que $f : G \rightarrow Y, Y \in \{0, 1\}$, Y correspond à l'étiquette binaire.

La particularité de la modélisation réside dans la constitution des arcs. Les informations se propagent alors via le processus de convolution au sein du modèle final. L'hypothèse proposée est que des événements proches dans le temps ont des propriétés comparables. Cela se traduit dans le graphe par le calcul d'une proximité temporelle p , qui sert de poids aux arcs, pour chaque couple (e_i, e_j) à partir de leurs coordonnées temporelles $T = \{t_1, t_2, \dots, t_n\}$. Ainsi, plus deux événements sont proches dans le temps, plus le poids de l'arc qui les relie est important. La décroissance progressive du poids au fur et à mesure de l'éloignement temporel

ressemble à la notion d’oubli étudiée en sciences cognitives. Nous avons comparé plusieurs fonctions d’oubli par expérimentation et retenu la fonction développée dans White (2001) Π , qui est définie par :

$$\Pi(b, \delta_{i,j}) = \exp(-b * \delta_{i,j}) * (\delta_{i,j} \geq 0), b > 0, \delta_{i,j} = t_i - t_j \quad (1)$$

δ correspond à la distance temporelle calculée sur chaque couple d’évènements. b est un paramètre de la modélisation, qui influe sur l’intensité de la décroissance de la fonction. La valeur de ce paramètre a un effet direct sur la structure des graphes générés.

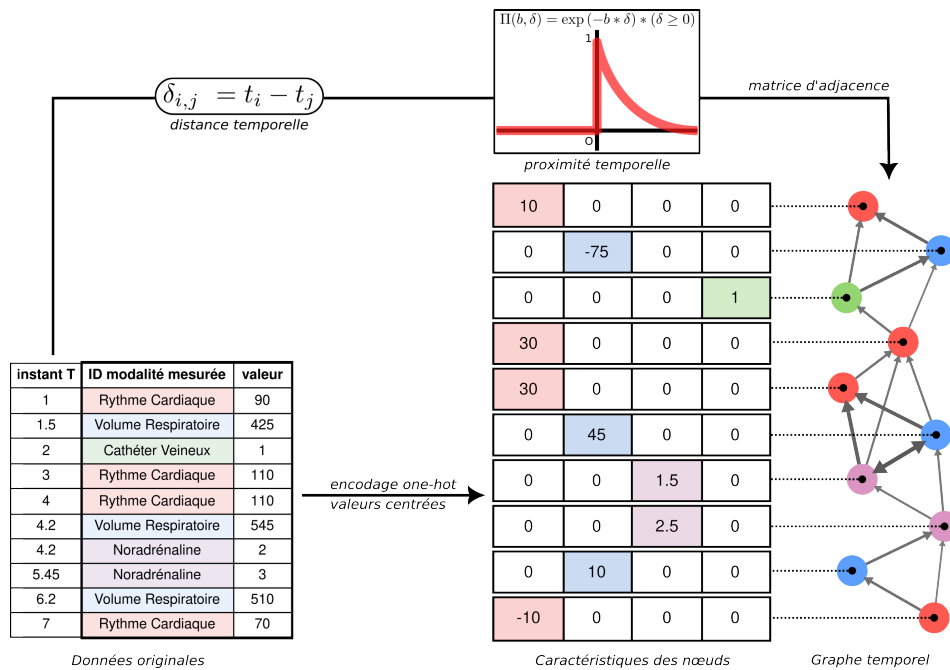


FIG. 1 – Processus de transformation des données originales vers un graphe d’évènements. Illustration issue de données fictives.

La Figure 1 récapitule les différentes étapes de la modélisation proposée, qui part des données pour produire un graphe de proximité temporel : d’un côté, la valeur centrée et la modalité de chaque mesure sont transformées dans une matrice où chaque colonne correspond à une modalité mesurée. Chaque ligne correspond à la représentation vectorielle d’un nœud donné. De l’autre côté, les coordonnées temporelles sont transformées en distance par paire d’évènements pour être transformées en proximité par l’application de la fonction Π (voir l’équation 1). Le résultat est une matrice d’adjacence qui décrit le poids des arcs du graphe final.

3 Expérimentations

Pour évaluer notre approche, nous nous sommes intéressés à la tâche de prédiction du décès à 24h décrite dans Harutyunyan et al. (2019)¹. Seuls les parcours patients qui durent au moins 4h sont sélectionnés. Pour chacun d’entre eux, nous cherchons à prédire toutes les heures si un patient décède dans les 24h suivantes, à partir des données antérieures à l’observation. Cette tâche a donc l’avantage de s’appuyer sur des séquences de tailles variées.

Les données originales sont enregistrées sur Physionet (Goldberger et al., 2000) dans la base publique MIMIC-III Clinical Database (Johnson et al., 2016b,a). Cette base est souvent utilisée dans la littérature pour comparer les approches pour de nombreuses tâches de supervision médicale car elle est volumineuse, hétérogène et déjà anonymisée. Les données mises à disposition décrivent des parcours de patients réels provenant de l’activité du Beth Israel Deaconess Medical Center de 2001 à 2021. Pour leurs expérimentations, Harutyunyan et al. (2019) sélectionnent 17 descripteurs selon des critères médicaux. Les valeurs sont filtrées, prétraitées et associées à des graduations numériques dans le cas de données catégorielles ordonnées. Les volumes de données sont décrits dans la Table 1. Afin d’obtenir une comparaison équitable avec nos propres expérimentations, nous nous appuyons exactement sur ces données transformées, dont toutes les étapes d’obtention sont disponibles publiquement.

Nombre de patients du jeu d’entraînement	28 620
Nombre de patients du jeu de test	5 058
Pourcentage de patients de la classe minoritaire	11,0%
Durée moyenne des séjours	86.6h
Durée maximale des séjours	2 103h

TAB. 1 – *Caractéristiques des données extraites de MIMIC-III.*

Notre implémentation utilise les couches de convolution de graphes issues de la bibliothèque DGL (DeepGraph Library). Le modèle présenté est choisi après la recherche des hyperparamètres et de la structure optimales grâce à la bibliothèque Optuna (Akiba et al., 2019). Les résultats présentés pour l’application de notre méthode sont issus d’une validation croisée à 5 blocs, découpés dans le jeu d’entraînement puis évalués sur le jeu de test.

Le modèle désigné par “Convolution de Graphes” dans la Table 2 consiste en une succession des couches suivantes : trois couches de GraphSAGE (Hamilton et al., 2017) avec la moyenne pour fonction d’agrégation, puis la concaténation de trois fonctions d’agrégation *min*, *max* et *moyenne* sur la représentation de tous les nœuds (*readout_nodes*), deux couches linéaires et un softmax en sortie. Une fonction d’activation *tanh* est placée après chaque couche d’apprentissage. L’optimiseur choisi est Adam avec le taux d’apprentissage fixé à 0,003 et la décroissance des poids à 0,05. Le paramètre de proximité temporelle *b* est fixé à 4,5, la largeur des couches successives est fixée à 15. L’implémentation du calcul des proximités temporelles définie dans la Section 2 sous forme parallélisée est limitée par la mé-

1. Plus précisément, Harutyunyan et al. (2019) présente 4 tâches dont la prédiction de la décompensation. Il s’agit d’une dégradation, souvent brutale, d’un organe ou d’un organisme qui était jusqu’alors maintenu en équilibre par des mécanismes de compensation qui empêchaient la survenue de ce dérèglement. Les auteurs utilisent alors le critère de décès pour sélectionner les patients à qui ils associent l’étiquette de décompensation.

moire vive disponible. Pour cette raison, seuls les 5 000 derniers évènements d'un parcours sont retenus. L'apprentissage est réalisé sur une machine dotée de 40 cœurs de calculs, 126Go de RAM et 4 cartes graphiques NVIDIA GeForce GTX 1080, comportant 11Go de VRAM chacune.

Modèle	AUC
Harutyunyan et al. (2019) :	
Régression Logistique	0.870
LSTM - S	0.892
LSTM - C	0.906
LSTM - C + DS	0.911
Notre méthode :	
Convolution de Graphes	0.897

TAB. 2 – Comparaison des résultats sur la prédiction de décès dans les 24h. Les différentes versions du LSTM sont extraites du papier original avec pour significations respectives “S” : Standard, “C” : Channel-wise et “DS” : Deep Supervision.

Pour comparer les approches, nous mesurons l'aire sous la courbe sensibilité/spécificité, aussi désignée comme AUC, sur le jeu de test. C'est un score calculé à partir de la courbe qui donne le taux de vrais positifs en fonction du taux de faux positifs, en faisant varier le seuil de décision de la classification binaire. Ce score est adapté aux jeux de données dont les proportions d'étiquettes sont déséquilibrées. Cette approche est comparée aux résultats des méthodes utilisées dans Harutyunyan et al. (2019), à savoir la régression logistique et les différentes variantes de LSTM.

Comme nous pouvons le constater dans la Table 2, notre modélisation offre des résultats similaires à ceux de l'état de l'art, observés dans Harutyunyan et al. (2019). Notre objectif, dans cet article, est de valider la représentation des données patients sous la forme d'un graphe. Le résultat obtenu par notre méthode dépasse certaines versions de LSTM.

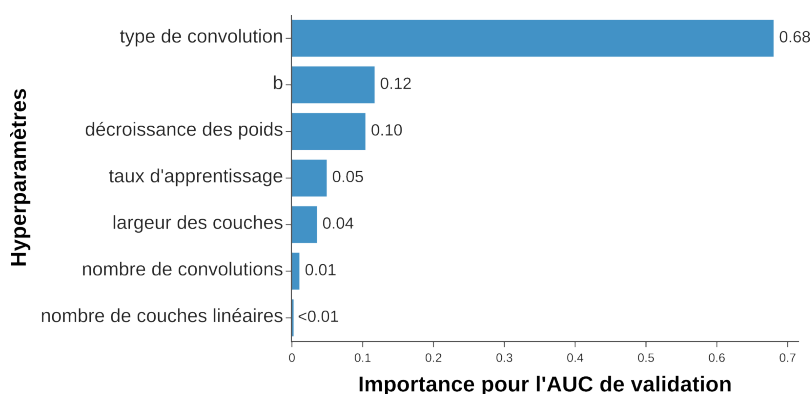


FIG. 2 – Impact des hyperparamètres sur les performances à l'issue de l'optimisation par Optuna.

On peut se demander à quel point les prédictions du modèle présenté sont influencées par la transmission d'informations des GNN à travers les arcs issus de la modélisation comme attendu plutôt que par d'autres effets. La Table 2 montre la valeur contributive des différents hyperparamètres durant l'optimisation des hyperparamètres par Optuna. Les résultats sont issus d'une analyse fonctionnelle de la variance, i.e. fANOVA (Hutter et al., 2014). On observe que les hyperparamètres les plus importants sont le type de convolution utilisée et le paramètre b de la modélisation. On peut en déduire que la constitution des arcs et la manière dont les informations se propagent au sein du graphe dans le modèle ont un fort impact sur les performances mesurées.

4 Conclusions et perspectives

Dans cet article, nous avons proposé une nouvelle modélisation des données patients sous la forme de graphe. Nous utilisons les arcs pondérés pour représenter la proximité temporelle entre deux événements. Nous avons expérimenté cette modélisation en tirant parti des convolutions de graphes et montré que nos résultats approchent ceux de l'état de l'art. Nous proposons donc une représentation alternative des données pour la prise en compte du temps. Du fait de la modularité offerte par les structures de graphes et des nombreux apports récents dans le domaine des GNN, il est possible d'imaginer divers axes d'améliorations.

La distance δ en entrée de la fonction de proximité ne prend en compte que l'horodatage de début de l'évènement. De nombreuses mesures stockées dans la base ont une durée donnée. Il faudrait donc adapter la distance temporelle à des intervalles plutôt qu'à des coordonnées ponctuelles. Par exemple, les données démographiques seraient ainsi considérées comme des éléments de contexte visibles pendant toute la durée du parcours. En général, cela permettrait une meilleure description des mesures continues et de leurs superpositions dans le temps.

Nous allons expérimenter d'autres protocoles d'apprentissage. Actuellement, la tâche de prédiction consiste à prédire la valeur de l'étiquette par rapport au graphe entier. Or, d'après Dwivedi et al. (2020), les réseaux de neurones appliqués aux graphes, obtiennent des résultats significativement meilleurs que des réseaux classiques pour des tâches comme la prédiction de liens ou la classification de nœuds alors que les gains de performance restent limités pour la classification de graphe. Nous allons donc considérer le décès non plus comme une étiquette exogène au graphe mais comme un nœud à inclure dans le graphe. Nous allons alors prédire soit l'existence du nouveau nœud étiquette, soit celle d'un arc entre un tel nœud et les autres mesures du patient. Le principal avantage de cette approche est que si le modèle n'a besoin que d'informations intrinsèques pour la supervision, alors tous les nœuds du graphe peuvent être considérés comme étiquettes d'intérêt. Il s'agirait d'apprendre les relations entre événements en auto-supervision, sans avoir recours à une étiquette extérieure et indépendamment des tâches à résoudre. Avec cette stratégie, on peut bénéficier de l'intégralité des données sans recourir à un étiquetage manuel, qui s'avère souvent très coûteux. Cela aurait pour effet de mutualiser le coût de l'apprentissage à travers des utilisations finales variées.

Outre la recherche de l'amélioration des performances de prédiction, la valeur explicative d'un modèle est un aspect important pour son adoption par les professionnels de santé. Nous visons à développer des extensions qui peuvent apporter des aspects complémentaires en plus des performances pures comme l'explicabilité, l'interprétabilité ou la découverte de connaissances. Le choix du graphe permet la projection visuelle de la structure de données dans un

espace visuel en deux dimensions (Herman et al., 2000). De plus, on observe l'apparition de méthodes complémentaires qui permettent d'extraire des informations des GNN dans le but d'étayer les prédictions. Par exemple, Ying et al. (2019) simplifient le graphe d'entrée, en sélectionnant les nœuds et arcs qui contribuent le plus à la prédiction. On peut imaginer intégrer ce type de méthodes pour mettre en valeur les éléments contributifs du graphe dans un système visuel destiné aux soignants.

Références

- Akiba, T., S. Sano, T. Yanase, T. Ohta, et M. Koyama (2019). Optuna : A next-generation hyperparameter optimization framework. In *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Ashfaq, A., A. Sant'Anna, M. Lingman, et S. Nowaczyk (2019). Readmission prediction using deep learning on electronic health records. *Journal of Biomedical Informatics* 97, 103256.
- Dwivedi, V. P., C. K. Joshi, A. T. Luu, T. Laurent, Y. Bengio, et X. Bresson (2020). Benchmarking graph neural networks. *arXiv*.
- El-Rashidy, N., S. El-Sappagh, T. Abuhmed, S. Abdelrazek, et H. M. El-Bakry (2020). Intensive care unit mortality prediction: An improved patient-specific stacking ensemble model. *IEEE Access* 8, 133541–133564.
- Goldberger, A. L., L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, et H. E. Stanley (2000). PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* 101(23), E215–220. Place: United States.
- Hamilton, W., Z. Ying, et J. Leskovec (2017). Inductive representation learning on large graphs. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, et R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, Volume 30. Curran Associates, Inc.
- Harutyunyan, H., H. Khachatrian, D. C. Kale, G. Ver Steeg, et A. Galstyan (2019). Multitask learning and benchmarking with clinical time series data. *Scientific Data* 6(1), 96.
- Herman, I., G. Melancon, et M. Marshall (2000). Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics* 6(1), 24–43.
- Hutter, F., H. Hoos, et K. Leyton-Brown (2014). An efficient approach for assessing hyperparameter importance. In *Proceedings of International Conference on Machine Learning 2014 (ICML 2014)*, pp. 754–762.
- Johnson, A., T. Pollard, et R. Mark (2016a). MIMIC-III Clinical Database. Version Number: 1.4 Type: dataset.
- Johnson, A. E. W., T. J. Pollard, L. Shen, L.-W. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, et R. G. Mark (2016b). MIMIC-III, a freely accessible critical care database. *Scientific Data* 3, 160035.
- Jung, B., A. Daurat, A. De Jong, G. Chanques, M. Mahul, M. Monnin, N. Molinari, et S. Jaber (2016). Rapid response team and hospital mortality in hospitalized patients. *Intensive Care*

Modélisation de parcours patients : graphes temporels pour la supervision médicale

Med. 42(4), 494–504.

Li, Y., S. Rao, J. R. A. Solares, A. Hassaine, R. Ramakrishnan, D. Canoy, Y. Zhu, K. Rahimi, et G. Salimi-Khorshidi (2020). Behrt: Transformer for electronic health records. *Scientific Reports* 10(1), 7155.

Poongodi, T., D. Sumathi, P. Suresh, et B. Balusamy (2021). Deep learning techniques for electronic health record (EHR) analysis. In A. K. Bhoi, P. K. Mallick, C.-M. Liu, et V. E. Balas (Eds.), *Bio-inspired Neurocomputing*, Studies in Computational Intelligence, pp. 73–103. Springer.

White, K. G. (2001). Forgetting functions. *Animal Learning & Behavior* 29(3), 193–207.

Ying, R., D. Bourgeois, J. You, M. Zitnik, et J. Leskovec (2019). GNN explainer: A tool for post-hoc explanation of graph neural networks. *CoRR abs/1903.03894*.

Zhou, J., G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, et M. Sun (2020). Graph neural networks: A review of methods and applications. *AI Open* 1, 57–81.

Remerciements

Ce projet est financé par une bourse CIFRE, financée par 5 DEGRÉS, établie en collaboration avec le LIRMM et le CHU de Montpellier.

Summary

Machine learning methods are becoming increasingly popular to anticipate critical risks in patients under surveillance thus reducing the burden on caregivers. In this paper, we propose an original modelling that benefits of recent developments in Graph Convolutional Networks: a patient’s journey is seen as a graph, where each node is an event and temporal proximities are represented by weighted directed edges. We evaluated this model to predict death at 24 hours on a real dataset and successfully compared our results with the state of the art.