Graph neural networks and causal inference for wireless networks management

PhD study proposal (CIFRE)

Summary

The objective of this thesis is twofold: the theoretical exploration of the connections between graph neural networks (GNNs) and causal inference methods; the practical development of critical mobile networks applications based on these tools. There are two main applications we target in this scope that are currently addressed by GNN: radio resource management (RRM) and predictive maintenance (PM) of wireless networks. The first application requires fast and scalable algorithms which generalize to new use-cases in 6G and future wireless systems. The second application requires accurate and explainable predictions, which help to identify the root cause of faults. We believe that causal inference offers a formal framework which makes it possible to generalize and improve on our current work.

RRM problems are inherently based on an underlying graph, representing the wireless network, whose properties can be efficiently learned by GNNs. GNNs are promising to tackle problems that are currently intractable due to their high computational complexity or lack of mathematical model. There is a myriad of such problems including UE-AP association, power control, precoding, scheduling in frequency and time, at the transmitter side; and at the receiver side: channel estimation and active user detection. We aim at developing a meta-learning GNN framework generalizing to a large class of resource allocation problems, constraints, objective functions. Causal inference techniques will help to *fuse* together data arising from different environments and data distributions. One of the purposes of this thesis is to extend the GNN framework from an offline to an online setting with data-fusion utilizing both simulated data generated offline and online data collected from different systems.

As the complexity and the size of telecommunication networks increases, maintenance becomes an increasingly challenging task, demanding intelligent algorithms for decision-making. Our team has recently developed a data-driven approach to detection and prediction of faults in the hardware components of base stations, leveraging the alarm logs produced by the network elements. There are two natural graph structures that we can leverage on our data: the topology of the base stations, and the correlation structure between different alarm types. GNNs can incorporate both these graph structures in a predictive model. Recent internal work has already shown this approach's potential. Although the above-mentioned graph structures can be interpreted causally, the GNN model lacks explainability, which is crucial for this kind of application. To this end, a further goal of this thesis is providing a causal interpretation to the GNN model.

.

Description

In this section we provide more details about the relevant applications which motivate the need to bring together GNNs and causal reasoning.

RRM applications

In wireless communications, radio resource management (RRM) consists in allocating the physical resources (frequency, time, power, etc.) between pairs of transmitter-receiver devices to optimize some desired metrics such as spectral efficiency, age of information, power consumption or energy efficiency. An important metric is the signal-to-interference-plus-noise-ratio (SINR) which indicates the quality of a received signal. A device's SINR value has a direct impact on the data rate received by the device. These problems usually fall into the category of constrained optimization problems. They can be convex or non-convex, continuous or discrete, deterministic or stochastic, centralized or distributed, offline or online.

For some problems, the state-of-the-art solutions (e.g., optimal algorithms) cannot be computed in practice due to their high computational complexity. For this reason, we develop low complexity approximations that are implementable in real systems. GNNs are promising tools as they can efficiently learn the non-Euclidean properties and invariances of wireless communication graphs. Furthermore, their time complexity is linear in the graph size, which makes them massively scalable. An example of such model applied to RRM is the random edge GNN [Eisen 2020] which outperforms classical deep learning models such as multilayer perceptron (MLP) and convolutional neural network (CNN). The authors of [Shen 2023] discuss the benefits and limitations of GNNs for wireless problems.

Recent work in Bell Labs [Salaün 2022] shows that GNN reaches unprecedented performance and generalizability for precoding and power control problems in 6G and beyond wireless networks. This work targeted Cell-Free Massive MIMO - a key 6G wireless technology. We believe that GNNs could be applied to other resource allocation problems in an analogous way. However, developing a GNN optimized for a single task in a specific setting, as commonly done in literature, limits its applicability. Indeed, such models must often be reworked when deployed in a system with divergent constraints. For this reason, we adopt a meta-learning perspective. The goal is to design a meta-model that learns the representation shared by several RRM tasks. This meta-model can then be specialized (trained on specific data) and deployed seamlessly on a wide range of use-cases in various environments, with different system requirements and configurations.

Once deployed, the meta-model may need to adapt to the real environment quickly, with few observations. The data distribution in the real-case scenario can differ considerably from training data. Recent work has shown that few-shot learning and speed of adaptation improve if we leverage the invariances across environments. If we use the correct representation variables and know the causal mechanisms describing their interaction, changes in distribution across environments are sparse and controlled, increasing the speed of adaptation as a result [Bengio 2019].

A typical input of RRM algorithms is the channel state information (CSI) which characterizes the signal propagation in the environment. When the channel is known within its coherence time, we refer to instantaneous CSI. In contrast, statistical CSI refers to the channel statistics over a longer timescale than the channel coherence time. In practice the CSI is estimated and can thus be noisy or delayed depending on the wireless system’s capabilities and requirements.

RRM machine learning models can be either offline or online. Offline methods are trained outside the wireless system before deployment. When deployed, they are executed independently of any past/future CSI and allocation. These schemes can work with both perfect and imperfect (noisy) CSI, but do not consider the temporal effect of the environment. We also refer to these methods as one-shot. Online algorithms allocate resources over-time given potentially time-varying and noisy CSI. The online setting can consider user mobility, changing environment and objective, delayed or missing CSI at certain times. An online model can continuously adapt to these changes, while an offline model must be frequently retrained to avoid model drift issues between training and deployment. However, online models have poor performance when initially deployed and usually require a large amount of data before reaching acceptable performance. Moreover, training in a real system is significantly more costly than in a simulated environment.

Integrating the advantages of offline and online methods in a unique data-driven approach is a challenging but crucial task. Since online and offline data rely on distinct representations of channel state information, the associated data distributions cannot be naively identified. Causal analysis [Pearl 2009, Bareinboim 2016] provides a mathematical framework for formalizing this data-fusion problem. Online and offline data share part of the underlying structural causal relationships, and this invariance can be leveraged to transfer knowledge from the offline to the online case. Similar techniques can also be applied to transfer data and models across different systems, sharing common causal mechanisms. The goal is to explore the applicability of causal inference techniques to data fusion and implement machine learning models for RRM which can learn simultaneously from online and offline data.

Predictive Maintenance applications

As the complexity and size of telecommunication networks increases, maintenance becomes an increasingly challenging task, demanding intelligent algorithms for decision-making. Our team has recently approached this predictive maintenance task in a data-driven way, leveraging the thousands of alarm logs produced daily by the mobile network base stations [Massaro 2023]. The alarms data carries two natural graph structures: we know which device issues the alarm and its location in the network and base station topology (spatial proximity); we know when each device issue an alarm and when alarms are firing together (temporal proximity). Both these graph structures can be leveraged and incorporated in the predictive model. Ongoing research has shown that GNNs can encode the network and devices topology into the predictive model, leading to an improvement in faults prediction and showing the potential of this approach. For instance, predicting the fault of a device/node in a network can be rephrased as a node classification task. The GNN structural bias is crucial, as the anomaly of one device could be indicated by anomalies on the neighboring devices.

Both the spatial and temporal graph structures are directed acyclic graphs, which can be interpreted causally (e.g., Granger causality for the temporal graph). Yet the GNN model lacks a causal interpretation, which would be useful for explaining the outcome of a prediction. If the GNN model entails a causal model, it could be used to estimate causal effects and to provide robust root cause interpretation of faults. Some seminal work in this direction has been recently spelled out [Zečević 2021]. In our setting, we can assume that the behavior of each base station is characterized by a few causal mechanisms depending on the aggregation of parent's nodes in the causal graph, hence reflecting the standard GNN structural bias. On the theoretical level, the goal is to generalize the standard notion of structural causal model [Pearl 2009] to this setting. On the practical level, apply the developed causal inference tools to the predictive maintenance task.

References

[Pearl 2009] Judea Pearl. *Causality*. Cambridge university press, 2009.

[Bareinboim 2016] Elias Bareinboim, Judea Pearl. *Causal inference and the data-fusion problem*. PNAS. vol. 113, no. 27, 2016.

[Bengio 2019] Yoshua Bengio et al. *A Meta-Transfer Objective for Learning to Disentangle Causal Mechanisms.* ICLR 2020*.*

[Eisen 2020] Mark Eisen, Alejandro Ribeir. *Optimal wireless resource allocation with random edge graph neural networks*. IEEE Transactions on Signal Processing, vol. 68, pp. 2977-2991, 2020.

[Zečević 2021] Zečević et al., *Relating Graph Neural Networks to Structural Causal Models*, arXiv 2109.04173, 2021.

[Salaün 2022] Lou Salaün et al. *A GNN approach for cell-free massive MIMO*. IEEE Global Communications Conference. 2022.

[Massaro 2023] Massaro et al. *Predicting Network Hardware Faults through Layered Treatment of Alarms Logs*. Entropy, vol. 25, no. 6, 2023.

[Shen 2023] Yifei Shen et al. *Graph neural networks for wireless communications: from theory to practice*. IEEE Transactions on Wireless Communications, vol. 22, no. 5, pp. 3554-3569, 2023.

Contacts

Lou Salaün, lou.salaun@nokia-bell-labs.com, Bell Labs, Massy, France

Davide Rinaldi, davide.rinaldi@nokia-bell-labs.com, Bell Labs, Massy, France