THÈSE DE DOCTORAT DE

NANTES UNIVERSITÉ

ÉCOLE DOCTORALE N° 641
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : *Computer Science*

Par

Sarah BENIKHLEF

Federated learning of Bayesian networks preserving privacy for personalised medical applications

Thèse présentée et soutenue à Polytech Nantes, le 10/01/2025
Unité de recherche : Laboratoire des Sciences du Numérique de Nantes (LS2N - équipe Duke)

Rapporteurs avant soutenance :

Nahla BEN AMOR	Professeur ISG de Tunis
Christophe GONZALES	Professeur Aix Marseille Université

Composition du Jury :

Attention, en cas d'absence d'un des membres du Jury le jour de la soutenance, la composition du jury doit être revue pour s'assurer qu'elle est conforme et devra être répercutée sur la couverture de thèse

Président :	Prénom NOM	Fonction et établissement d'exercice *(à préciser après la soutenance)*
Examinateurs :	Karim TABIA	Professeur Université d'Artois
	Hala SKAF MOLLI	Professeur Université de Nantes
Dir. de thèse :	Philippe LERAY	Professeur Université de Nantes
Co-dir. de thèse :	Guillaume RASCHIA	Maître de conférence Université de Nantes

COLLEGE ED MATHEMATIQUES
DOCTORAL STIC
PAYS DE LA LOIRE

Nantes Université

Titre : Apprentissage fédéré de réseaux bayésiens préservant la confidentialité dans le cadre d'applications médicales personnalisées

Mot clés : Réseaux bayésiens, apprentissage multi-tâches, apprentissage par transfert, confidentialité différentielle

Résumé : L'apprentissage fédéré permet d'entraîner des modèles d'apprentissage automatique sur plusieurs ensembles de données décentralisés sans partager les données brutes, répondant ainsi aux préoccupations critiques en matière de confidentialité dans le domaine de la santé. Les réseaux bayésiens (RB) sont des modèles probabilistes qui ont prouvé leur valeur dans la modélisation de dépendances complexes. Ils ont la particularité d'être grandement interprétables, ce qui est crucial lors du traitement de données médicales, car cela permet aux experts médicaux de comprendre facilement les modèles construits. Cette thèse étudie l'intégration des réseaux bayésiens dans un cadre d'apprentissage fédéré, en mettant l'accent sur la préservation de la confidentialité. Une première proposition combine les réseaux bayésiens avec des préoccupations statistiques en apprentissage fédéré telles que l'apprentissage multi-tâches et l'apprentissage par transfert. Elle consiste en une extension d'un algorithme efficace de découverte de structure RB, MMHC, à un contexte multi-tâches (MT). Pour évaluer cette approche, nous proposons une procédure pour générer des benchmarks MT à partir de n'importe quel modèle de référence. Enfin, nous considérons les contraintes de confidentialité différentielle pour rendre l'algorithme d'apprentissage des RBs respectueux de la vie privée.

Title: Federated Learning of Bayesian Networks preserving privacy for personalised medical applications

Keywords: Bayesian networks, multi-task learning, transfer learning, differential privacy

Abstract: Federated learning allows machine learning model to be trained across multiple decentralized datasets without sharing raw data, addressing critical privacy concerns in healthcare. Bayesian networks (BN) are probabilistic models that have proven their worth in modeling complex dependencies. They have the particular ability to be highly interpretable, which can be crucial when dealing with medical data, as it allows medical experts to easily comprehend the constructed models.

This thesis investigates the integration of Bayesian networks within a federated learning framework, with a focus on privacy preservation. A first proposition combines Bayesian networks with statistical concerns in federated learning such as multi-task learning and transfer learning. It consists on an extension of an efficient BN structure discovery algorithm MMHC to a multi-task (MT) context. To evaluate this approach, we propose one procedure to generate MT benchmarks from any reference model. Finally, we consider differential privacy constraints into to make the BN learning algorithm privacy-conscious.