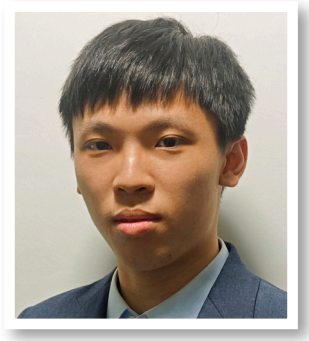
RESEARCH

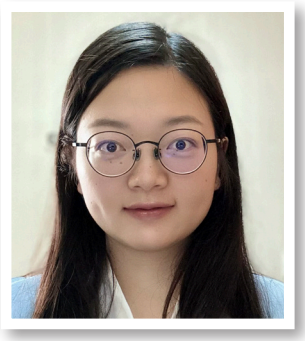
Inside IISE Journals

This month, we spotlight two articles from the upcoming special issue on Federated Distributed Learning and Analytics to appear in *IISE Transactions* (Vol. 57, No. 7). These articles showcase how federated learning advances information exchange and causal discovery across distributed industrial systems while preserving data privacy. The first article presents a novel data analytics method for discovering nonlinear causal relationships across distributed systems while preserving data privacy. Addressing the limitations of traditional Bayesian networks in causal analysis, the authors propose a two-step, federated multitask learning approach that enables each system to model its local nonlinear causal structures within its database. Meanwhile, a central server aligns these structures collaboratively without accessing raw data to discover global causal structures. The method demonstrates high accuracy while ensuring strong privacy protection. This research offers a powerful solution for collaborative causal discovery in privacy-sensitive domains such as distributed manufacturing systems. The second article addresses the challenge of secure information exchange for collaborative data analysis among organizations while preserving individual privacy. The proposed algorithms are built on differential privacy, enabling the addition of controlled noise to the shared data to ensure confidentiality. Validated using real-world energy consumption data from U.S. power grids and German households, as well as clinical trial data from multiple healthcare centers, the approach demonstrates improved accuracy in data analytics. This research represents a significant advancement in privacy-preserving collaborative learning, providing a robust solution for data analytics when direct data sharing is restricted.

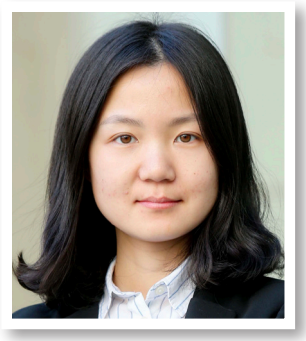


Tian Lan

Ben Niu



Xing Yang



Chen Zhang

Learning causality across systems, together and apart: A federated thinking

Understanding not just what is correlated, but why things happen – uncovering true cause-effect relationships across variables – is essential for understanding system behavior and enabling informed decision-making. Bayesian networks, which represent variable connections through directed acyclic graphs, have emerged as powerful tools for modeling their causal effects. However, most traditional Bayesian networks typically assume linear dependencies between variables, which may oversimplify the complex, nonlinear dynamics present in many real-world systems.

In real-world industrial environments, relevant data are often distributed across multiple systems or organizations, each capturing different yet related variables under varying conditions. Although combining these datasets could significantly improve causal analysis, privacy, regulatory and proprietary concerns often render

direct data sharing infeasible. Federated learning has emerged as a solution for integrating diverse datasets while preserving data privacy. The challenge involves discovering causal relationships across systems with both overlapping and distinct variables without compromising data confidentiality.

To address these challenges, researchers Xing Yang and Ben Niu from Shenzhen University, together with Tian Lan and Chen Zhang from Tsinghua University, propose a new approach that integrates nonparametric causal modeling with federated multitask learning. Their method enables the discovery of nonlinear causal relationships across distributed datasets – without exposing raw data.

In this approach, each system models its local Bayesian network with smooth basis functions that flexibly capture nonlinear causal patterns within its own dataset. A central server then coordinates the learning process by aligning causal structural similarities across systems. This two-step optimization process facilitates the collaborative discovery of global causal insights while maintaining tailored models for each system's specific variable set and ensuring strict data privacy.

The proposed approach demonstrates superior performance in both accuracy and data privacy protection, as evidenced by simulations and a real-world case study of a three-phase flow facility. This work presents a powerful approach for collaborative, privacy-preserving causal discovery across complex, distributed systems.

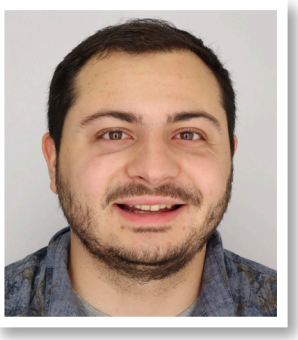
CONTACT: Chen Zhang, zhangchen01@tsinghua.edu.cn; Department of Industrial Engineering, Tsinghua University, Beijing, China, 100084

Privacy-preserving information exchange for collective inference and learning

How can companies share and analyze data collectively while complying with privacy regulations? Can smart grids, healthcare systems and financial institutions exchange information without violating their obligations to protect the privacy of their customers, patients, clients, and other individuals? How can we effectively reconcile individual privacy and data security requirements with the goals of collective decision-making and learning?

Understanding the tradeoffs between data privacy and utility in collective learning scenarios is critical to designing efficient learning systems that balance accurate decision making and individual privacy.

In their *IISE Transactions* paper, "Privacy-Preserving Distributed Estimation and Learning," Marios Papachristou, Ph.D. candidate at Cornell University, and Amin Rahimian, assistant professor of Industrial



Marios Papachristou

Amin Rahimian

Engineering at the University of Pittsburgh, emphasize the relevance of these challenges to the energy sector, where emerging net metering technologies in smart grids need to support distributed energy generation; for example, to compensate customers for rooftop solar while protecting consumer data. Sharing detailed energy consumption data in this case can pose significant privacy and security risks, such as revealing when someone is at home, their daily commute, online habits, family illnesses and so on.

The authors have developed algorithms to estimate statistical properties of private signals that are distributed across a network of agents. For example, agents may represent different households, signals may reflect net metering measurements, and the goal is to estimate the average power consumption for optimal pricing. Their approach builds on the paradigm of differential privacy, a widely used framework that protects individual data by adding controlled noise to shared information. Their designed algorithms can adjust noise levels based on the variability of the private signals and the communication patterns among agents.

To test their approach, they applied it to real-world energy consumption datasets collected from power grids in the U.S. and households in Germany. They demonstrated that their methods not only maintain privacy but also produce accurate results efficiently, outperforming conventional approaches that use differential privacy with federated learning.

In a follow-up work, "Differentially Private Distributed Inference," Papachristou and Rahimian study similar tradeoffs for inference in discrete spaces, particularly for distributed hypothesis testing such as in a multicenter clinical trial. In that case, multiple health centers collaborating on a clinical trial face privacy risks for their patient data. Without reliable privacy guarantees, their data sharing and analytics would require complex legal agreements to comply with healthcare privacy regulations, such as HIPAA.

CONTACT: Amin Rahimian, rahimian@pitt.edu; (267) 393-2376; Department of Industrial Engineering; University of Pittsburgh, 3700 O'Hara Street, Pittsburgh, PA 15261