LASAR^3
Laboratory of Analysis of Systems for the Assessment of Reliability, Risk and Resilience

Multi-Agent Framework for Learning Causal Bayesian Network Structures

Context of the research

Bayesian Networks (BNs) are widely used for the risk assessment and management of systems. Their key advantage is the ability to represent causal relationships among variables, enabling interpretable reasoning, scenario analysis, and decision support. However, learning BN structures from data alone remains a major challenge, as current structure-learning algorithms struggle to reliably distinguish causality from correlations, especially when data are limited, noisy, or partially observed.

In real-world engineering applications, many sources of system-specific and domain knowledge exist: expert understanding of causal mechanisms, system descriptions, and technical documentation. Yet, such information is often unstructured, heterogeneous, and difficult to formalize into priors or structural constraints for BN learning.

Recent advances in Large Language Models (LLMs), particularly when combined with Retrieval-Augmented Generation (RAG) and the capability of using computational tools (e.g., dedicated software and coding libraries), offer the capabilities for extracting and using domain knowledge. These models can integrate structured and unstructured information, interact with users, directly modify the BN structure and validate it with dedicated accuracy and causality metrics.

The research proposed here aims to leverage this new generation of Artificial Intelligence (AI) systems by constructing a multi-agent framework combining LLM reasoning, RAG, expert knowledge, BN construction and validation tools to generate interpretable and accurate causal BN structures.

The methodology will be tested on benchmark datasets for causal discovery and on a realistic case study such as a cyber–physical system (CPS) (e.g., power grid integrated with telecommunication network and control center).

Objective of the research

This research aims to develop a multi-agent framework to generate interpretable, accurate causal BN structures for the risk assessment and management of complex systems. The activity will include:

- Literature review on LLMs and their application in safety engineering;
- Definition of the research problem;
- Development of the methodology;
- Application to benchmarks and realistic case studies;
- Validation;
- Writing of scientific papers, as chapters of the thesis.

Required competencies and skills

- Interest in developing innovative applications of Natural Language Processing, Machine Learning and Artificial Intelligence algorithms for reliability, availability, maintainability and safety of complex system;
- Good knowledge of Python programming or willingness to learn.

Composition of the research group

- 2 full professors;
- 1 postdoctoral researchers;
- 1 PhD student.

For further information, please contact:
Prof. Piero Baraldi, piero.baraldi@polimi.it
Prof. Enrico Zio, enrico.zio@polimi.it