Towards Bayesian Symbolic Computational Graph Completion

A Roadmap for Scientific Knowledge Infusion into Symbolic Machine Learning
Tim Schneider M.Sc. Prof. Dr. Wolfgang Nowak Prof. Dr. Steffen Staab
ARTIFICIAL INTELLIGENCE SOFTWARE ACADEMY (AISA) KICKOFF 29.04.2022

Recent progress in Machine Learning is great, but . . .

Current Machine Learning (ML) is often driven by large amounts of low quality internet data, yielding large, computationally expensive, black box models that provide little insight into the problem. When directly applied to Scientific Domains, such methods . . . require a large amount of data to converge to a good solution. In many scientific domains only little data is available yielding to overfitting and unsatisfactory results.

To address these problems . . .

1. Encode Prior Knowledge in a (Bayesian) Graph

Black Box Solutions . . . provide black box solutions and therefore do not create new scientific hypotheses. But without the latter, ML only adds little value to the scientific domain research.

Example: 1D Lake Temperature Model

Task: find the unknown function \( f(x, d) \rightarrow T \)

"I know that temperature \( T \) is related to density \( \rho \) via \( g \) and the density monotony \( h \) holds with increasing depth \( d \)" (compare [1])

Our Approach

- Encode the problem as well as all prior knowledge in a computational (factor) graph representation.
- Fill in symbolic solutions with data and Bayesian Reasoning.

Scientific Knowledge Exploitation

Hereby, the framework considers three steps that exploit the prior knowledge in the graph:

(1) Unnormalized Posterior: use the graph to evaluate the unnormalized posterior distribution of potential symbolic expressions in order to sample from it using MCMC (see [2]).

(2) Informed Sampler: involutive maps [5] encode additional information about the problem to design informed samplers that produce samples of high likelihood. Especially learned maps [6] allow a connection to recent advances in deep learning (e.g. [7]).

(3) Graph Update: The Bayesian solver yields symbolic hypothesis that enables a human researcher to accept one of the hypothesis / candidates or come up with new prior knowledge in the graph.

References