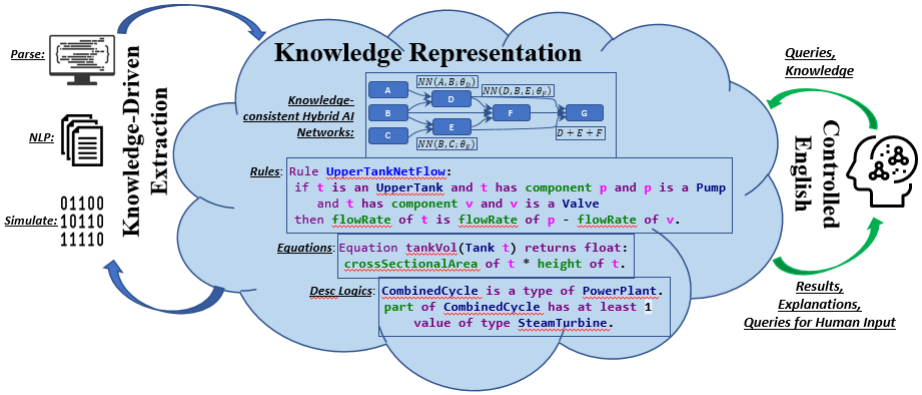
Augmented Bayesian **N**etworks Integrating **S**emantics **W**ith **E**xtraction and **R**eadability (ANSWER)

GE Global Research; Andrew W Crapo, PhD

DARPA-PA-18-02
Artificial Intelligence Exploration (AIE)

CONCEPT



APPROACH

- Capture scientific knowledge in a modular Bayesian architecture that combines Description Logics-style semantics capturing domain knowledge with equations for physics-based models and constraints and neural networks for data-driven models and imprecise knowledge including causality.
- Create a novel local-global Bayesian inference over all types of knowledge to enable representation accuracy, efficient convergence, task-specific performance, and reusability.
- Extract scientific models from code via parse tree analysis and simulation and from documentation and publications via custom named entity recognition and application of domain knowledge .
- Use controlled-English grammars to facilitate query formulation, make knowledge readable and explorable, and explain results of inference and model evaluation and execution.

IMPACT

Need: To understand and extract scientific knowledge, expressed in code, documentation, or publications, requires deep contextual knowledge ranging from domain concepts to the laws of physics to the syntax and semantics of the source. Extracted knowledge is of various types but must be usable as an integrated whole with inference and querying capability yielding results easily readable and understandable by humans.

Goal: Significantly advance the state-of-the-art by 1) combining disparate knowledge types into a single cohesive Bayesian knowledge network capable of understanding and using scientific models, 2) extracting scientific models from code and text via a knowledge-driven and focused approach, and 3) enabling high-bandwidth bidirectional communication and collaboration between a knowledgeable AI and a human.

CONTEXT

Current approaches:

- Graphical models make use of dependencies but cannot depict exact equations and constraints.
- Bayesian calibration with physics equation and experimental data can learn parameters and capture unmodeled dynamics but do not integrate knowledge of causality and constraints.
- Code analysis tools focus on software errors and vulnerabilities and do not maximally exploit parse trees to extract high-fidelity scientific models.
- NLP is used to do named entity extraction but is not sufficiently informed by domain knowledge and unique "fingerprints" of in-text equations and scientific knowledge.

Source Selection Information