SAND 2011-7622 C
SAND2011-7622C

Exceptional service in the national interest

Sandia National Laboratories

A Plausible Logic Inference Engine

David J. Stracuzzi

November 4, 2011

U.S. DEPARTMENT OF ENERGY
NNSA National Nuclear Security Administration
LDRD LABORATORY DIRECTED RESEARCH & DEVELOPMENT

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

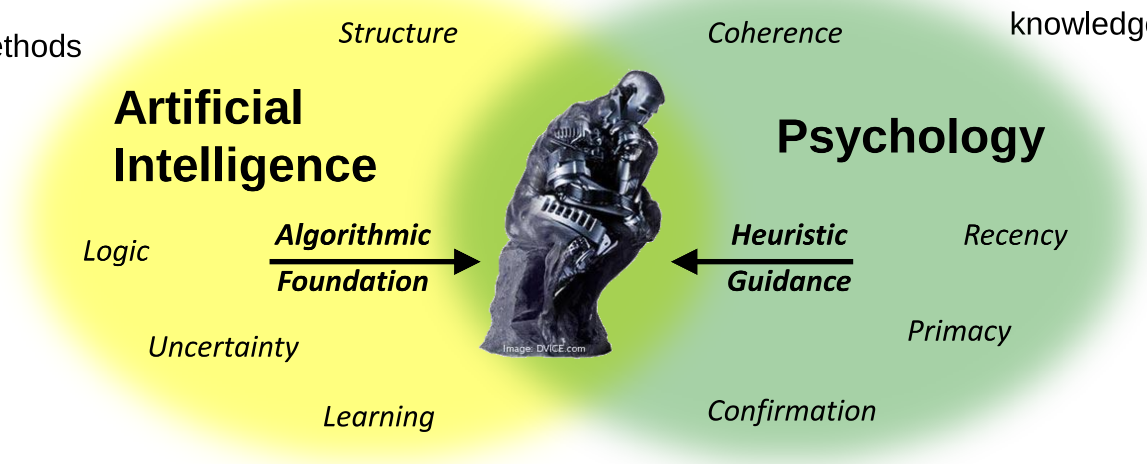
Overview

Background: Reasoning serves as a foundation for many high-level cognitive capabilities.

- Many AI approaches
  - Statistics and optimality
  - Logic and proof
- Several psychological theories
  - Concerned with source of error vs. source of power
  - Lacking in detail
- Both study different methods in isolation

Goals: Develop psychologically plausible methods for computational reasoning that reflect the strengths of human reasoning

- Flexible: nature and content of data
- Computationally efficient
- Pragmatic: useful absent optimality
- Self-sufficient: minimal reliance on knowledge engineers



Problem: Modern methods cannot emulate the flex-ibility, efficiency, or pragmatism of human reasoning.

- Reliant on brute-force computation
- Emphasis on proof and optimality
- Little connection to psychology/cognition literature

Approach:

- Start with a cognitive systems view
- Combine multiple reasoning techniques
- Drop notions of optimality and provability
- Incorporate known psychological biases
- Integrate learning ability

Knowledge Representation

The system has a foundation of first-order logic.

Predicates: represent conditions of the environment

$$Neighbor(X,Y)$$

Beliefs:

- Represent specific instances of a condition
- Stored in *working memory*
- Generated by inference or direct observation
- Include plausibility score

$$Neighbor(\mathrm{BOB}, \mathrm{PETE})^{\mathrm{Pl}=E^{+}-E^{-}}_{1.0=1.0-0.0}$$

$$-1 \le \mathrm{Pl}(b) \le 1; \quad \mathrm{Pl}(b) = -\mathrm{Pl}(\neg b)$$

Rules:

- Represent generalized relationships among predicates
- Stored in *long-term memory*
- Use threshold logic
- Annotated with reliability factors

$$(v_0 + v_1 \cdot \mathrm{Pl}(Alarm(X)) + v_2 \cdot \mathrm{Pl}(Neighbor(X,Y)) > 0)$$

$$Alarm(X) \wedge Neighbor(X,Y) \underset{\alpha=0.9}{\Rightarrow} Calls(Y,X)$$

$$\underset{\beta=0.1}{\Rightarrow} (w_0 + w_1 \cdot \mathrm{Pl}(Calls(Y,X)) > 0)$$

$$0 \le \alpha, \beta \le 1$$

Key idea: Working memory structures are activations of long-term structures; consistent with Cowan's (1988) view that WM is an extension of LTM.

Inference Patterns and Evidence Propagation

Key idea: Logically unsound inference patterns are still useful for accumulating evidence about beliefs (Polya, 1954)

Example: Accepting the Consequent (Abduction)

Given a rule

$$(v_0 + v_1 \cdot \mathrm{Pl}(Alarm(X)) + v_2 \cdot \mathrm{Pl}(Neighbor(X,Y)) > 0) \xRightarrow[\beta=0.1]{\alpha=0.9} (w_0 + w_1 \cdot \mathrm{Pl}(Calls(Y,X)) > 0)$$

And evidence

$$Neighbor(\mathrm{BOB, PETE})^{\mathrm{Pl}=1.0}, \quad Calls(\mathrm{PETE, BOB})^{\mathrm{Pl}=1.0}$$

Draw a conclusion

$$v_0 + v_1 \cdot \mathrm{Pl}(Alarm(X)) + v_2 \cdot \mathrm{Pl}(Neighbor(X,Y)) > 0$$

Generate a new belief

$$Alarm(\mathrm{BOB})$$

Compute change in evidence

$$\Delta E^{+}(Alarm(\mathrm{BOB})) = \beta \frac{\sum_{j \in S} w_j \mathrm{Pl}(Q_j)}{\sum_{j \in S} |w_j|} \approx 0.1$$

Summary: Inference patterns govern the mechanics of combining beliefs with rules to derive new beliefs. Other patterns are similar.

Heuristics

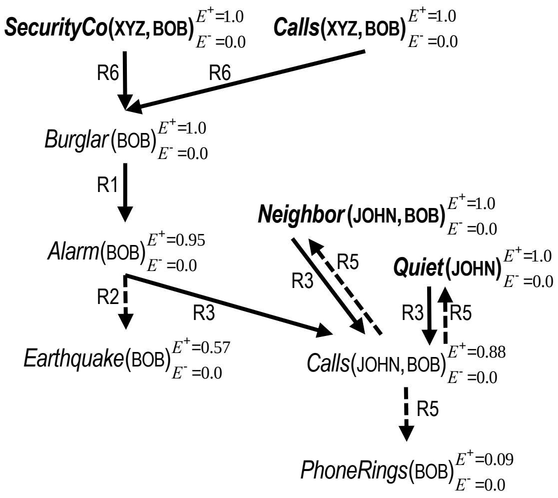
Key idea: Many inferences are possible; apply known psychological biases as guidance heuristics.

Primacy & Recency (evidence update):

$$E_{t+1}(Alarm(\mathrm{BOB})) = (1- / v_1)E_t(Alarm(\mathrm{BOB})) + / v_1 \Delta E_t(Alarm(\mathrm{BOB}))$$
$$\approx 0.03$$

Coherence (working memory):

$$\mathrm{coh}(b) = \sum_{b' \in \mathrm{adj}(b)} \frac{w_{b'}\ \mathrm{rel}(b', b)}{\delta(b')+1} \mathrm{Pl}(b')$$



Confirmation (plausibilities):

- Prefers rule applications with stronger evidence over weaker
- Ignores low-plausibility beliefs

Summary: Biases based on properties of beliefs. Aimed at reasoning along a small number of plausible trajectories.

Related and Future Work

Related work:

- Abductive Rationalizing Agents – Bridewell and Langley (2011)
- Extended Plausible Inference – Friedman (1981)
- Production Systems
- Markov Logic – Richardson and Domingos (2006)
- Bayesian Logic Programs – Kersting and DeRaedt (2005)

Future work:

- Agent goals – Adding and using goal information as a source of bias
- Analogical reasoning – Identifying and constructing analogies
- Learning – Acquiring and modifying rules and analogical mappings
- Metacognition – Adding to and modifying the inference patterns

Questions and Comments?

Extra Slides

Application: Plausible Reasoning to Support Analyst Decision Making

STATE OF THE ART

- Analysts review incoming data for “interesting situations”
- Increasingly overwhelmed by growth in data volume
- Modern computational inference techniques cannot emulate the flexibility, efficiency, or pragmatism of human reasoning
  - Computationally expensive
  - Questionable scalability
  - No connection to psychological literature or human capability

NEW INSIGHTS

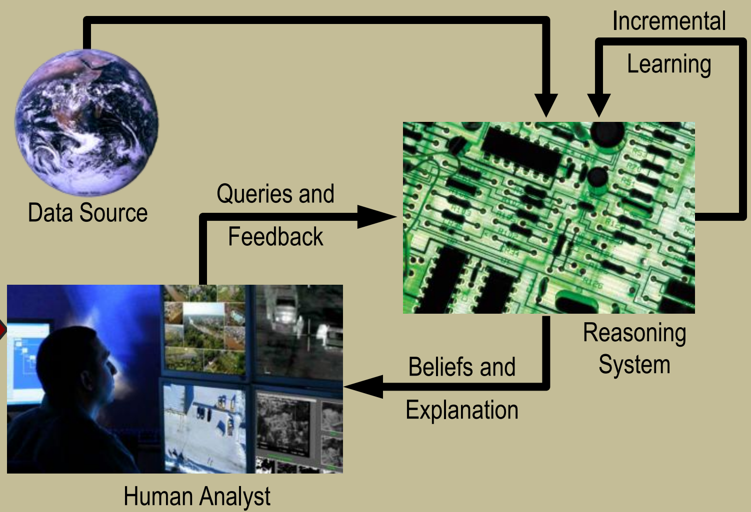
Integrate complimentary results from artificial intelligence and psychology to produce a flexible, efficient, and practical system.

- Combine logical, statistical, evidential reasoning techniques (each brings needed capability to bear)
- Integrate psychological biases (provide guidance to inference process)
- Integrate learning with inference (improve knowledge with experience)
- Keep the human in the loop (contrast to much work in AI)

Collaborative Data Analysis

Human analyst and computational reasoning system collaborate to:

- *Assess and explain data*
- *Expand knowledge about task and domain*
- *Improve speed and accuracy of assessments*
- *Improve outcomes associated with resulting decisions*



Reasoner provides computational power, but also constructs a model of analyst knowledge. This allows efficient, high level interactions, and provides a way to review analyst knowledge and methods.

GOALS

Create a psychologically plausible computational reasoning system that reflects the strengths of human reasoning and can help to develop insights into the fundamental mechanisms that guide human reasoning.

- Flexible (nature and content of data/information)
- Efficient (computation to produce results)
- Pragmatic (useful results absent optimality)

IMPACT

Approach leverages computational speed and human creative power to analyze large quantities of data.

Applications include:

- Situation assessment
- Cyber security
- Knowledge capture
- Foundation for realistic agent control (e.g. training simulators)

Current State of the Art

Artificial Intelligence

- Dominance of statistical inference and learning methods
  (e.g. Bayes Nets and various generalizations, SAT-based methods)
  - Data intensive
  - Computationally expensive
  - Questionable scalability
- Logic-based approaches more efficient, but brittle
  (e.g. rule engines, inductive logic)
  - Rely heavily on knowledge engineering
  - Do not capture uncertainty
  - Limited reasoning capacity
- Little or no connection to psychological literature

Psychology

- Multiple theories of human reasoning method and ability
- Frequently driven by attempts at accounting for errors
- Models tend to be theoretical and high-level
- Few detailed computational models of human reasoning

System Overview

The learning engine monitors the stream of observations, beliefs, and rules to discover new or modify existing rules.

Key Points:

- Applies empirical evidence to expand and add nuance to long-term memory
- Engine runs in the background
- Changes occur incrementally

…ing memory contains *beliefs,* or specific …ed in

Pattern memory contai… …plied during inference, analogy, and… Examples: *Modus Pone…* *…ntecedent*

Key Points:

- Each pattern contai… …lic part, and a statistical/numeri…
- Computationally si… …constructing new beliefs, associa…
- Incremental incorporation of evidence (plausibility update)

Key Point…

- Simpl…
- Logic softened with degree of belief in the rule
- Nuanced with empirical experience

Observations

Working Memory

Inference Engine

Analogy Engine

Learning Engine

Long-Term Memory

Pattern Memory