Explainable AI via Argumentation: Theory & Practice

Antonis Kakas
University of Cyprus, Cyprus
Nikos Spanoudakis
Technical University of Crete, Greece

Co-founders (with P. Moraitis) of Argument Theory
https://www.argument-theory.com/

ESSAI 2024 School: 22-26 July, Athens

Lecture 5

- Summary of Course

- Hands-on Projects
  - Student Presentation of Projects.
  - Future development of Projects?

- Advanced Topics
  - Direct Authoring in Natural Language:
    - COGNICA with LLM
  - Explainable ML:
    - ArgEML: Learning Arg. Theories & Arg-Based Explainable Models Reading

SUMMARY

Argumentation-based Decision Making

- **Decision of option O:**
  - **Argument case supporting O – O is plausible**
  - **No argument case for any other O'**
    - **This is an IDEAL situation of optimal decision**
- **In PRACTICE, Good/Satisficing decisions.**
- **On-line on demand computation!**
- **Could be in Dilemma**
  - **A Difficult problem case ⇒ Need more information.**

Computational Argumentation: a "Roadmap"

From <Args, ATT> … to <Args, Att, Def> … to

… to <As, C, ב> …

… to GORGIAS <As U *Prs*, C > …

From Theory to Practice

… to *SoDA* Methodology for Knowledge Acquisition

… to rAIson

… to Applications

*So*ftware *D*evelopment via *A*rgumentation
SoDA Methodology

Identify the **Language** of **Options** & **Factors** for **Preference**

- **Consider application scenarios** and **state** the preferred/desired option(s) in each scenario.
  - **Identify** different **initial scenarios.**

- Successively **refine** the **scenarios**, **restating** at each **refinement** the new **preferred option(s).**

- Considering **combinations** of conflicting of **scenarios**

Hierarchies of Scenario-based Preferences (SBPs)

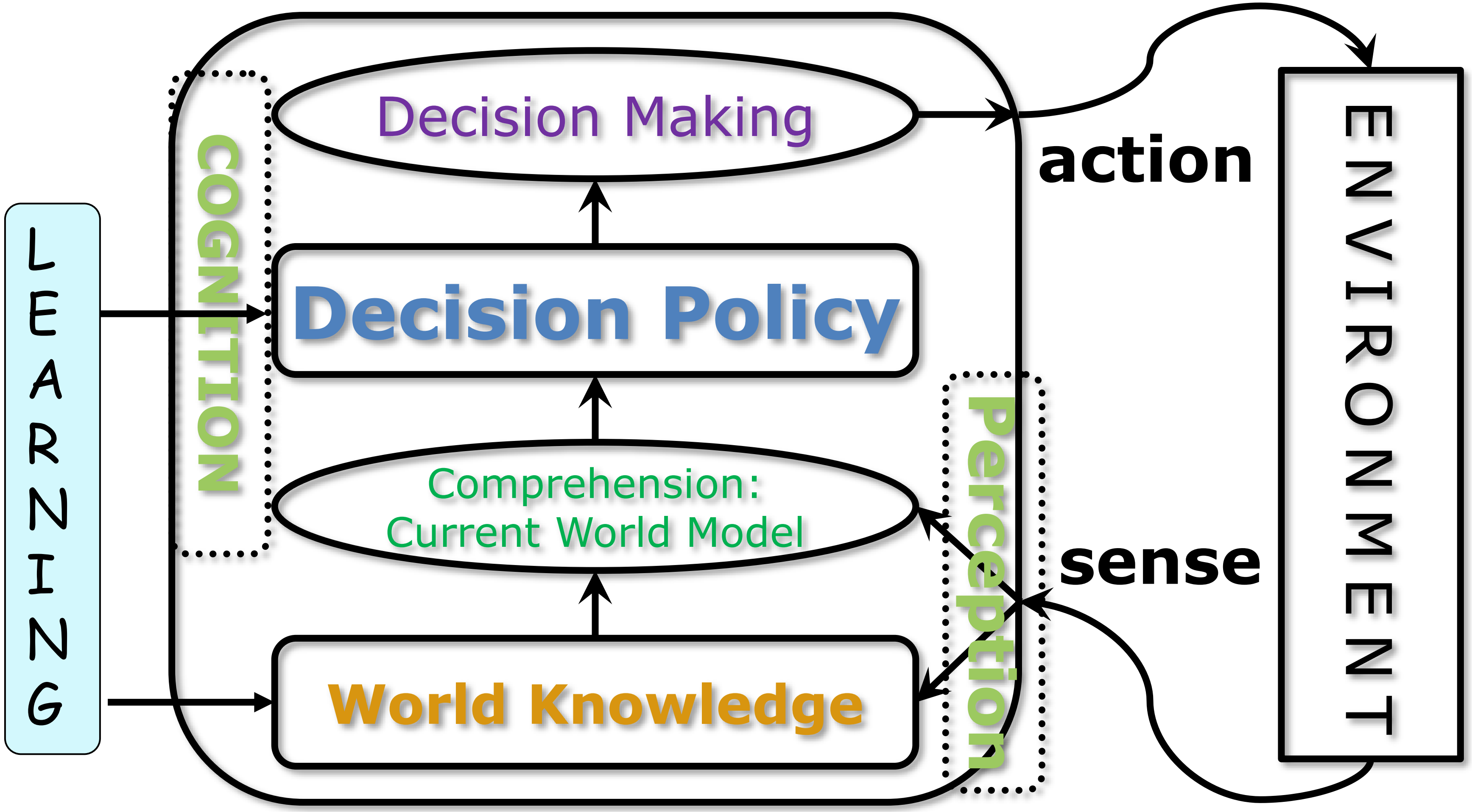
rAIson

Authoring - No coding
Knowledge Representation

Hierarchies of Scenario-based
Preferences (SBPs)

Executable
GORGIAS Argumentation Framework

Building Argumentation-based Decision Machines



Building Decision Machines

Two major Challenges

1. **Acquisition** of problem Knowledge - **Decision Policy**
   - **At a Language Level of the Application – Natural Language?**
   - **Extracting Hidden Preferences from Natural Language Specs**

Addressed by SoDA and rAIson

Building Decision Machines

Second major Challenge

2. **Middleware** from Sensory Information to Policy Concepts
   - **Comprehension** of current **Context** of the application environment from its low-level sensory information
   - In other words, **"Establishing the (high-level) facts."**

Intelligence is in the Abstraction of the Decision Policy
Large number of cases grouped into high-level concepts

Addressed by API of rAIson

Hands-on Projects

- Presentations/Discussion of Projects

- Future Development

- Redo a **more realistic** problem.
- **Finalize** your project.
  - Send us a report document
- **Keep** rAIson Accounts Open?
  - rAIson also suitable for Freelancers

ADVANCED TOPICS

COGNICA: Cognitive Argumentation

- **Cognitive Argumentation is a Synthesis of:**
  - **Argumentation theory in AI**
    - GORGIAS Argumentation framework
  - **Empirical and theoretical studies of the Psychology of Reasoning from Cognitive psychology and Philosophy**
    - P. N. Johnson-Laird and R. M. Byrne, "**Conditionals: a theory of meaning, pragmatics, and inference**." Psychol. Rev., vol. 109, (4), pp. 646, 2002.
- **Cognitive Machine Argumentation (2022) on ResearchGate**

COGNICAc

Decision Policy in Natural language

Policy complementary options

GPT 4 TURBO translates policy to Cognica Middle language (Cognica D.P. CNL)

Parse Cognica Middle language (Cognica D.P. CNL) to Cognica CNL.

Parse Cognica to GORGIAS Argumentation

Query Cognica on policy.

Get **Explanation** in **Natural Language**

COGNICAc Middle language

- **NORMALLY** {Conclusion}

- **IF** {Condition} **THEN** {Conclusion}
- **IF** {Condition} **THEN ALWAYS** {Conclusion}
- **IF** {Condition} **THEN MAYBE** {Conclusion}

- **IF** {Condition} **THEN** {Conclusion} **UNLESS** {Condition2} **IN SUCH CASES** {Conclusion2}

- **ONLY IF** {Condition} **THEN MAYBE** {Conclusion}
- **WHENEVER NOT** {Condition1} **THEN NOT** {Conclusion1} **BUT IF** {Condition2} **THEN MAYBE** {Conclusion1}

COGNICAc: Demo

Demo of Call Assistant

http://cognica.cs.ucy.ac.cy/COGNICAc/login.php

Building XAI Systems from Natural Specifications

Call Assistant (Personal Policy)

"Normally, allow a call. When at work deny a call from an unknown number. When busy at work deny a call unless it is an emergency family call. Always allow a call from my manager."

Options: allow a call, deny a call.

Busy at work

Busy at work, Emergency Family Call (not Unknown)

Busy at work, Manager call

Calendar Assistant in COGNICAc

"Business meetings should be scheduled in the afternoon unless it is with a very important customer. Also, when presenting final solutions to customers prefer a morning meeting.

Meetings with the CEO can be scheduled morning or afternoon. I prefer to have meetings related to ongoing internal projects in the morning.

Meetings with new customers must be scheduled in the morning."

Options: morning meeting, afternoon meeting.

ADVANCED TOPIC

ArgEML

Explainable Machine Learning via Argumentation

N. Prentzas, C. S. Pattichis, A. C. Kakas: ***Explainable Machine Learning via Argumentation****.* **xAI (3) 2023***: pp. 371-398.*

A. K. & L. Michael***, Abduction and Argumentation for Explainable Machine Learning: A Position Survey,*** https://arxiv.org/abs/2010.12896

Argumentation-based Machine Learning (ArgML)

Rules -→ Arguments

ML Prediction/Reasoning via Dialectic Argumentation

In ArgML Prediction can be: Definite or Dilemma

Argumentation-based Explainable Machine Learning *(ArgEML)*

ArgML ≡ *ArgEML*

Explanation Model ≡ Argumentation Theory

Argumentation-based Explainable Machine Learning *(ArgEML)*

Output of *ArgEML:*

A. Explanatory Predictions
Attributive and Contrastive Explanations

B. Explanatory Problem Partition (XPP)

ArgEML Metrics

- Definite Accuracy
- Amount of Ambiguity
- Quality of Explanations
- Compactness of Explanatory Problem Partition (XPP)

Medical Peer Companion Applications

Integrated XML with Symbolic Reasoning
Hybrid approach

- **CARDIOLOGICA: U-prevent Calculators to Reasoners**
- ***M. Sclerosis, Alzheimer, Aneurysm: XAI analysis***
- ***Emergency Hospitalization (ICU): Chart problem space***

Endometrial cancer Explanation Spaces

Compactness (Theory-b1) = 0.1/0.9 * 10 = 1.1

preferred

	Theory-b1				
Group	Explanation Pattern	prediction	type	coverage	accuracy
Group 1	r1	benign	definite	30%	79%
Group 2	r2	malignant	definite	27%	87%
Group 3	r3	benign	definite	11%	60%
Group 4	r4	malignant	definite	9%	67%
Group 5	p3;r1|p4;r9	dilemma	dilemma	3%	n/a
Group 6	p14;p1;r3|p2;r8	dilemma	dilemma	3%	n/a
Group 7	p14;r3|p15;r9	dilemma	dilemma	3%	n/a
Group 8	r3|r8	dilemma	dilemma	3%	n/a
Group 9	p5;r1|p6;r8	dilemma	dilemma	2%	n/a
Other				10%	

Compactness (Theory-b2) = 0.1/0.9 * 11 = 1.22

	Theory-b2				
Group	Explanation Pattern	prediction	type	coverage	accuracy
Group 1	r1	benign	definite	31%	80%
Group 2	r2	malignant	definite	27%	87%
Group 3	r3	benign	definite	11%	60%
Group 4	r4	malignant	definite	9%	67%
Group 5	p35;p14;p23;r3|p15;p36;p2;p34;r8	dilemma	dilemma	3%	n/a
Group 6	r3;p29;r1|p30;r8	dilemma	dilemma	3%	n/a
Group 7	p14;p35;r3|p2;p24;r9	dilemma	dilemma	2%	n/a
Group 8	p3;p25;r1	benign	definite	2%	71%
Group 9	p5;p27;r1|p2;p6;p28;r8	dilemma	dilemma	2%	n/a
Group 10	r3|p4;p26;r9	dilemma	dilemma	2%	n/a
Other				11%	

Aneurysm
Example Explanation

«This case is **Predicted to "Unrupture":**

This is **supported** by the fact "the aspect ratio is medium". This reason is **strengthened** against the reason "the mean curvature is medium" supporting **rupture** by the fact that the "mean radius is high" and the fact that the "location is not aca".

Also, the prediction for **urupture** in this case is **positively affected** by the facts: "size ratio is vlow" and "maneyrsm is 0". »

THANK YOU