Goal recognition, obfuscation-based decision making and cooperation

Nicolas Cointe nicolas.cointe@tudelft.nl
Amineh Ghorbani A.Ghorbani@tudelft.nl
Caspar Chorus C.G.Chorus@tudelft.nl

Delft University of Technology
Faculty of Technology, Policy and Management
Department of Engineering Systems and Services
Transport Policy and Logistics (TLO) section

European Research Council
Established by the European Commission

Our approach

Motivations

We want to design a mechanism to let agents evaluate the quantity of information given to an observer through a behavior. Then, we use this indicator in order to minimize or maximize her transparency.

Problem

How to measure and compare the transparency of agents' behaviors ?

Approach

We propose to embed in a classic Belief-Desire-Intention architecture a plan recognition (PR) algorithm to evaluate the probability of each plan of a plan library to be intended according with an observed behavior. Then the agent computes the Shannon entropy of this probabilities distribution as an obfuscation indicator.

Key concepts and applications

Transparency

A transparent agent makes decisions in order to both achieve her goals and make it as obvious as possible.

- increase legibility for human users
- contribute to the safety of a system
- useful to goal recognition-based coalition building

Obfuscation

An obfuscation agent is acting in order to both achieve a goal and hide it.

- Protect the privacy of a user
- Keep secret strategic information
- More realistic model for malicious agents

Interpreting a behavior

Perception

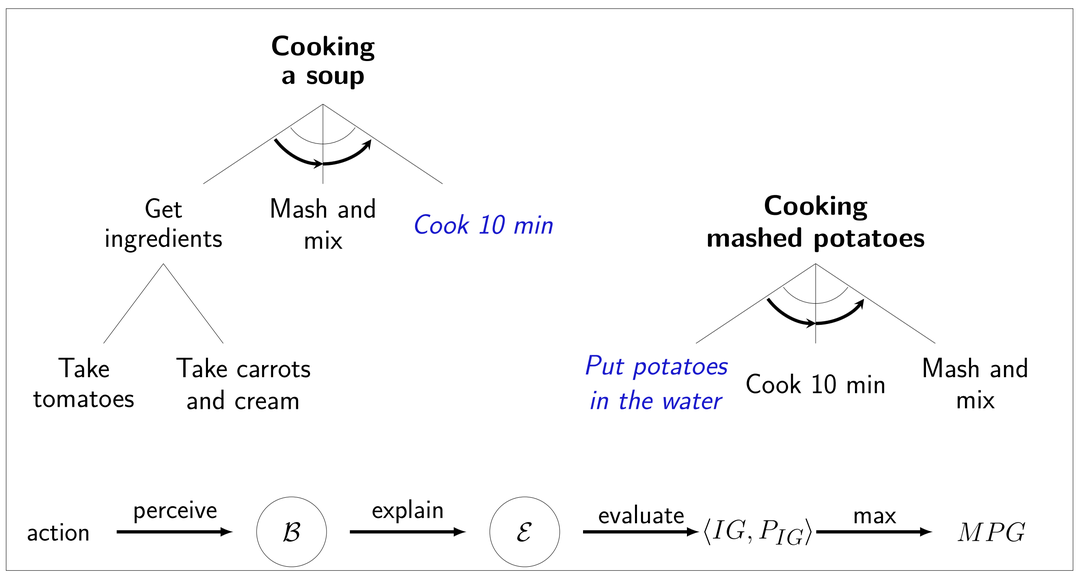
An *observed behavior* is an ordered set of actions executed by an agent and perceived during their execution in a shared environment.

Explanation

An *explanation* is set of observed actions associated with intendable goals. An explanation is *valid* if the plan library's constraints are respected. To face combinatorial explosion, the set of possible explanations is often not entirely explored.

Evaluation

The probability P_{ig} of an intendable goal according with an observed behavior and a plan library is given by a *plan recognition algorithm*. The more a root node is associated with actions in the set of the explanations, the higher the probability of this plan is.



Example

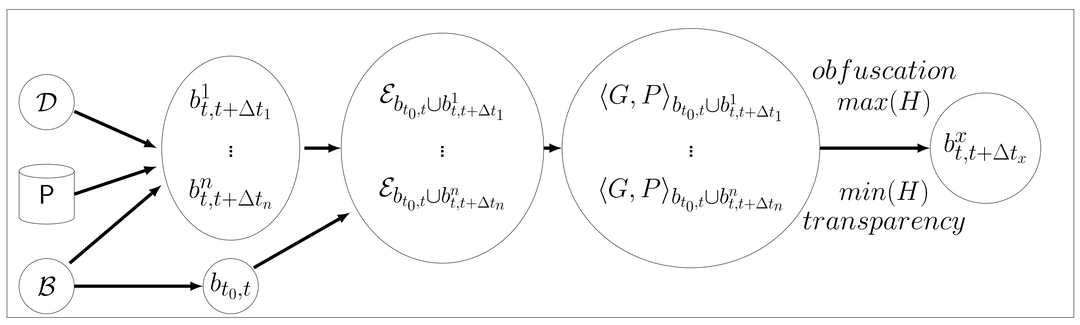
Observed behavior :
$b = \{$Put potatoes in the water, Cook 10 min $\}$

Explanations :
$\mathcal{E} = \{e_1, e_2\}$ with
e_1 = {{Put potatoes in the water, Cooking mashed potatoes}, {Cook 10 min, Cooking mashed potatoes}}
e_2 = {{Put potatoes in the water, Cooking mashed potatoes}, {Cook 10 min, Cooking a soup}}

As *Cooking mashed potatoes* is more often associated with the actions in the behavior
$P(Cooking\ mashed\ potatoes) > P(Cooking\ a\ soup)$

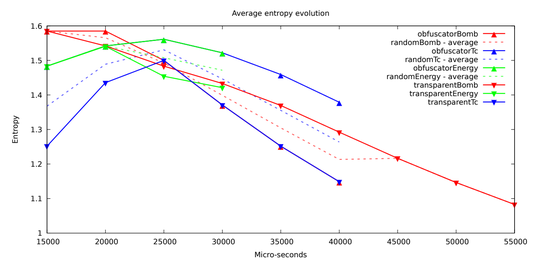
MPG =Cooking mashed potatoes

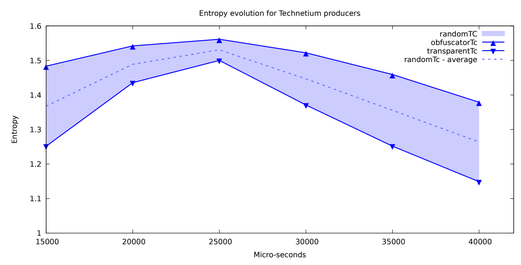
Obfuscation and transparency-based decision



When there is an OR-node in a plan, the agent should make a decision. To do so, she generates for each options the set of possible resulting behaviors. Then considering the past behavior and each hypothetical behavior, the decision making process infers the set of explanations for each possible decision. Finally, she computes the Shannon entropy on the entropy distribution for each decision. An obfuscator agent select then the decision associated with the highest entropy, and a transparent agent select the decision producing the lowest one.

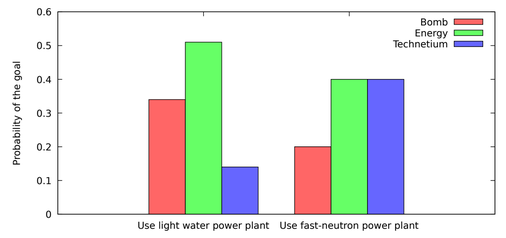
Experimental results





An implementation in JaCaMo to :

- Evaluate the impact of generated or manually described plan libraries (top left)
- Compare the behavior of different sets of agents, making random, obfuscation or transparency-based decisions (see top right)
- Visualize the probabilities distribution and entropy for each decision (see bottom right)



References

- Chorus C (2018) How to keep your AV on the right track ? An obfuscation-based model of decision-making by autonomous agents. In 7th Symposium of the European Association for Research in Transportation (hEART)
- Geib CW, Maraist J, Goldman RP (2008) A new probabilistic plan recognition algorithm based on string rewriting. In : ICAPS, pp 91–98
- Golman R, Hagmann D, Loewenstein G (2017) Information avoidance. Journal of Economic Literature 55(1) :96–135
- Mirsky R, Stern R, Gal K, Kalech M (2018) Sequential plan recognition : An iterative approach to disambiguating between hypotheses. Artificial Intelligence 260 :51–73
- Nyborg K (2011) I don't want to hear about it : Rational ignorance among duty-oriented consumers. Journal of Economic Behavior & Organization 79(3) :263–274

Perspectives

- Design a goal recognition-based trust to let the agent proactively decide who to collaborate with.
- Design a process to evaluate if another agent is more likely an obfuscator or a transparent agent.
- Apply it to a realistic scenario to show how the decisions make sense and how this model might be useful for the experts

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 724431).