Hintikka's World: scalable higher-order knowledge

Tristan Charrier[1*], **Sébastien Gamblin**[2], **Alexandre Niveau**[2,3] and **François Schwarzentruber**[4]
[1]First Affiliation
[2]Second Affiliation
[3]Third Affiliation
[4]Fourth Affiliation
{first, second}@example.com, third@other.example.com, fourth@example.com

Abstract

Hintikka's World is a tool that shows how artificial agents can reason about higher-order knowledge (agent a knows that agent b knows that...). In this demonstration paper, we present symbolic models that enables to implement in *Hintikka's World* large examples such as real card games.

1 Introduction

The current trend is to construct programs that play games with imperfect information, for instance Hanabi [Bard *et al.*, 2019], but also video games such as Starcraft 2 [Hu *et al.*, 2018] An important ingredient is to reason about higher-order knowledge (an agent knows that another agents knows that...). That is why we claim that epistemic logic and its dynamic extension, Dynamic epistemic logic ([Baltag *et al.*, 1998], [van Ditmarsch *et al.*, 2008]), may offer a formal tool to provide explanation in such AI programs. needs to be understood is relevant in AI, especially in strategic reasoning [Aumann, 1999].

The only tool we are aware of that enables to see and explore mental states of agents is *Hintikka's world* and was presented at ECAI-IJCAI 2018 [Schwarzentruber, 2018]. *Hintikka's world* is a proof of concept of a graphical user interface that represent Kripke models by comic strips, as shown in Figure 1. The tool is available at the following address: http://hintikkasworld.irisa.fr/.

Kripke models are graphs and they were represented explicitly in memory in the first version of the tool. Explicit models are useful to learn how dynamic epistemic logic works by means of toy examples: muddy children, Sally and Anne [Wimmer and Perner, 1983], etc. However, in real card games, such as Hanabi, there are .... possible configurations of cards. Thus, it is impossible to represent the full graph in memory: the first version of Hintikka's world does not *scale*.

That is why, we proposed to represent Kripke models symbolically by means of Binary Decision Diagrams as it was done in the tool DEMO[1] [van Benthem *et al.*, 2015].

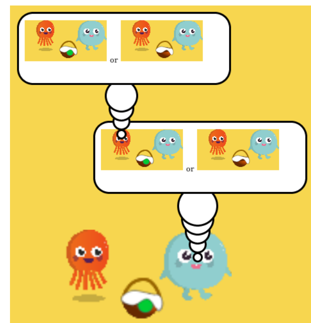

Figure 1: Graphical user interface of *Hintikka's world*

2 Symbolic models

TODO TIRSTAN

3 Demonstration Outline

TODO expliquer Hanabi et comment on peut y jouer

3.1 Already Implemented Examples

3.2 User Interaction

The tool adopts the point of view of Halpern and Vardi for modeling an epistemic situation: model checking is more suitable than theorem proving [Halpern and Vardi, 1991]. In other words, the current situation is modeled as a pointed Kripke model. By clicking on a given agent a, the interface opens a thought bubble that displays the possible worlds for agent a. Actually, the comic strips shows the unfolding of the current pointed Kripke model that represents the current situation.

On the left, the software shows buttons for possible actions (public announcement, public actions, private actions, etc.). Actions are modeled by pointed event models of Dynamic epistemic logic [Baltag *et al.*, 1998]. By clicking on a button, the corresponding action is executed: the product of the pointed Kripke model and the pointed event model becomes the current pointed Kripke model.

3.3 Building New Examples

The tool also allows the final user to building their own examples. They are two ways to specify a new epistemic situation.

*Contact Author

[1]The current implementation does not rely on DEMO since their work is not well-suited for a web application.

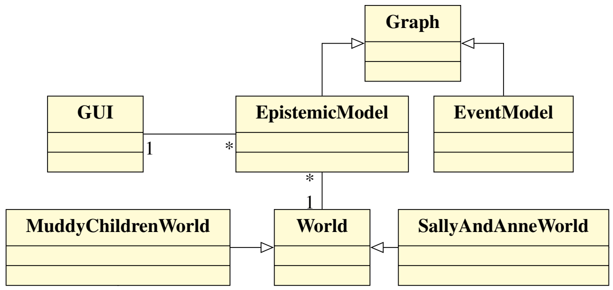

Figure 2: Architecture for the symbolic approach in *Hintikka's world*

First, the user can describe the initial pointed Kripke model in JavaScript, by giving the list of worlds, their valuations and the epistemic relations. Second, the user can specify the initial situation by a formula ϕ epistemic logic. The BNF is:

$$\begin{aligned}\phi := \;& p \mid (\texttt{not}\ \phi) \mid (\phi\ \texttt{and}\ \phi) \mid (\phi\ \texttt{or}\ \phi) \\ & \mid (\texttt{K a}\ \phi) \mid (\texttt{Kpos a}\ \phi) \mid (\texttt{CK G}\ \phi) \mid (\texttt{CKpos G}\ \phi)\end{aligned}$$

where p is an atomic proposition, a is an agent and G is a group of agents. E.g. 'p does not holds but agent a imagines that it is possible that p holds' (`((Kpos a p) and (not p)))`), agent a and b commonly know that agent c does not know the value of p (`(CK (a b) ((not (K b p)) and (not (K b (not p)))))`), etc. The user writes a set of formulas, one formula per line. Then the system solves the satisfiability problem and generates a pointed epistemic model.

4 System Description

TODO Expliquer comment c'est FAIT

4.1 Class Architecture

Figure 2 shows the main part of the architecture of *Hintikka's world*. The interesting part is the fact that the graphical user interface (GUI) is independent from the current example that is running (muddy children, Sally and Anne, etc.). In particular, adding a new example only requires to add a new class that inherits from `World` and to implement the method for drawing the scene from data (valuations, numbers, etc.) that are members of the class.

4.2 Model Checking

The tool highly rely on model checking. Indeed, for instance, performing the public announcement of ϕ requires to compute the subset of worlds in which ϕ holds and to prune the current Kripke model. We chose to write the model checking procedure in Javascript. Since model checking is in PTIME – thus is an easy task – and is used intensively, it suitable to run run it on the client-side for performance reasons.

4.3 Satisfiability Problem

5 Future Work

TODO implémzenter d'autres exemples etc.

References

[Aumann, 1999] Robert J. Aumann. Interactive epistemology I: knowledge. *Int. J. Game Theory*, 28(3):263–300, 1999.

[Baltag *et al.*, 1998] Alexandru Baltag, Lawrence S Moss, and Slawomir Solecki. The logic of public announcements, common knowledge, and private suspicions. In *Proceedings of the 7th conference on Theoretical aspects of rationality and knowledge*, pages 43–56. Morgan Kaufmann Publishers Inc., 1998.

[Bard *et al.*, 2019] Nolan Bard, Jakob N. Foerster, Sarath Chandar, Neil Burch, Marc Lanctot, Francis Song, Emilio Parisotto, Vincent Dumoulin, Subhodeep Moitra, Edward Hughes, Iain Dunning, Shibl Mourad, Hugo Larochelle, Marc G. Bellemare, and Michael Bowling. The hanabi challenge: A new frontier for AI research. *CoRR*, abs/1902.00506, 2019.

[Halpern and Vardi, 1991] Joseph Y. Halpern and Moshe Y. Vardi. Model checking vs. theorem proving: A manifesto. In *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning (KR'91). Cambridge, MA, USA, April 22-25, 1991.*, 1991.

[Hu *et al.*, 2018] Yue Hu, Juntao Li, Xi Li, Gang Pan, and Mingliang Xu. Knowledge-guided agent-tactic-aware learning for starcraft micromanagement. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden.*, pages 1471–1477, 2018.

[Schwarzentruber, 2018] François Schwarzentruber. Hintikka's world: Agents with higher-order knowledge. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden.*, pages 5859–5861, 2018.

[van Benthem *et al.*, 2015] Johan van Benthem, Jan van Eijck, Malvin Gattinger, and Kaile Su. Symbolic model checking for dynamic epistemic logic. In *Logic, Rationality, and Interaction - 5th International Workshop, LORI 2015 Taipei, Taiwan*, pages 366–378, 2015.

[van Ditmarsch *et al.*, 2008] Hans van Ditmarsch, Wiebe van der Hoek, and Barteld Kooi. *Dynamic Epistemic Logic*. Springer, Dordecht, 2008.

[Wimmer and Perner, 1983] Heinz Wimmer and Josef Perner. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1):103–128, 1983.