Reasoning About Muddy Children

Gregory Gelfond

AI Lab
Computer Science Department
Arizona State University

Overview

- ▷ Introduction
- ▷ The Muddy Children Problem
- ▷ Modal Logic
- ▷ Knowledge Actions
- ▷ Conclusion and Future Work

Introduction

In the area of multi-agent systems, it is often the case that agents must reason about the knowledge of other agents.

The muddy children domain is a classical problem which deals with this particular type of reasoning.

The Muddy Children Problem

Consider a domain in which we have n children playing. Some number of them, k, get mud on their foreheads. Each child can see the mud on the foreheads of the fellows, but not on themselves. Their father comes by and says "at least one of you has mud on their forehead." He then asks the following question repeatedly: "does any of you know whether you have mud on your own forehead?" Assuming that the children are perceptive, honest, and answer simultaneously, what happens?

It turns out that for the first $k-1$ times the father asks the question the children will say "no", but on the k^{th} time the children with muddy foreheads will all answer "yes". How can we capture this line of reasoning? The answer, at least in part, lies in the use of modal logic.

Modal Logic - Syntax

Suppose that we have n agents named $1 \ldots n$. We also introduce the *modal operators* $K_1, \ldots, K_n$.

Let Φ be a set of *primitive propositions*. We define a formula as follows:

- $\phi \in \Phi$ is a formula
- if φ is a formula, then $\forall i \in \{1 \ldots n\}$, $K_i\varphi$ is a formula; informally such a formula is read as "agent i knows φ"
- if φ is a formula, then $\neg\varphi$ is a formula
- if ψ and φ are formulas, then $\psi \wedge \varphi$, $\psi \vee \varphi$, $\psi \Rightarrow \varphi$, and $\psi \Leftrightarrow \varphi$ are formulas

Kripke Structures

Before we can proceed with defining the semantics of our modal logic, we must first introduce the notion of a Kripke structure, which provides us with a formalization of the notion of possible worlds.

A *Kripke structure* M for n agents over a set of primitive propositions Φ is a tuple $\langle S, \pi, R_1, \ldots, R_n \rangle$ where:

- $\triangleright$ S is a set whose elements are referred to as states or possible worlds
- $\triangleright$ π is an interpretation which associates with each state a truth assignment to the primitive propositions in Φ
- $\triangleright$ R_i is a binary equivalence relation (called an *accessiblity relation*) on S (i.e one that is reflexive, symmetric, and transitive)

For the purposes of this talk $(s_1, s_2) \in R_i$ if from the point of view of agent i the states are indistinguishable.

Note that unlike our standard conception of a state, the states of a Kripke structure are not completely defined by the interpretation of the primitive propositions comprising that state.

Modal Logic - Semantics

Given a Kripke structure M, a state s, and a formula φ, we define the entailment relation $(M, s) \models \varphi$ by induction on the structure of φ:

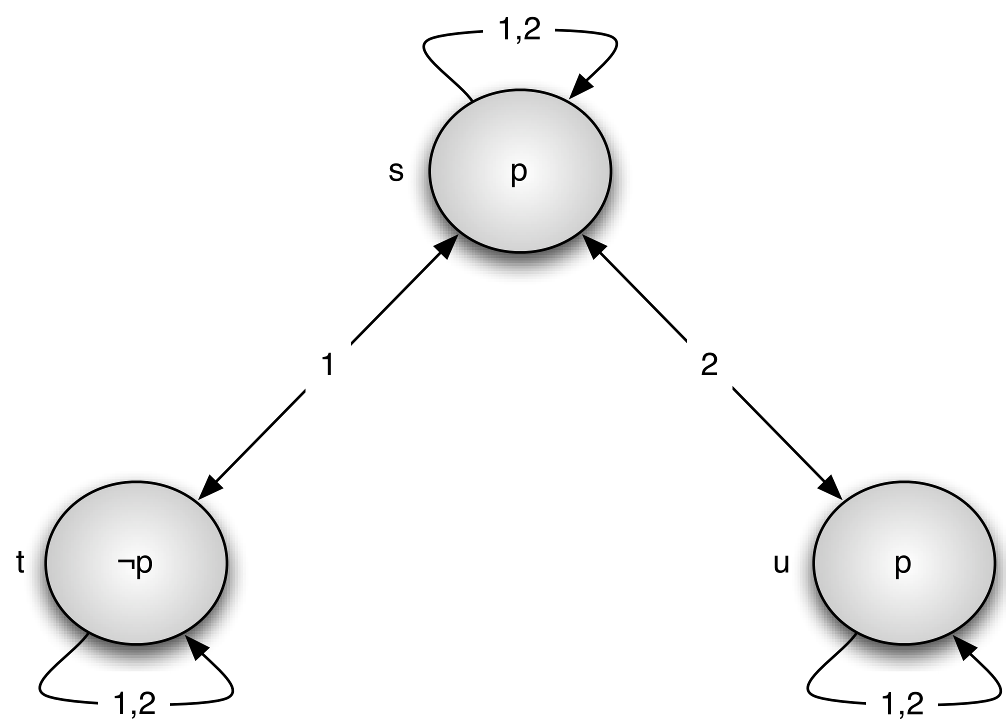
- $\forall \phi \in \Phi, (M, s) \models \phi$ if and only if $\pi(s)(\phi) = true$
- $(M, s) \models \neg\varphi$ if and only if $(M, s) \not\models \varphi$
- $(M, s) \models \varphi \wedge \psi$ if and only if $(M, s) \models \varphi$ and $(M, s) \models \psi$
- $(M, s) \models \varphi \vee \psi$ if and only if $(M, s) \models \varphi$ or $(M, s) \models \psi$
- $(M, s) \models K_i\varphi$ if and only if $(M, t) \models \varphi$ for all t such that $(s, t) \in R_i$

As $\psi \Rightarrow \varphi$ and $\psi \Leftrightarrow \varphi$ are used as shorthand for $\neg\psi \vee \varphi$ and $(\psi \Rightarrow \varphi) \wedge (\varphi \Rightarrow \psi)$ respectively, the entailment relation for these operators is not shown.

Finally, we say that $M \models \varphi$ if and only if for all $s \in S, (M, s) \models \varphi$.

Example

Consider the following Kripke structure M:



From the definition, we can see that:

- $(M, s) \models p$
- $(M, t) \not\models p$
- $(M, u) \models K_1 p \wedge K_2 p$
- $(M, s) \models K_2 p \wedge \neg K_1 p$
- $(M, s) \models K_1(K_2 p \vee K_2 \neg p)$

Common Knowledge

The modal operators K_i allow us to reason about the knowledge of a specific agent. In order to solve the muddy children problem however we need to be able to reason about the knowledge that is common to all of the agents of a particular group. This is accomplished by the introduction of some additional modal operators.

Let G be a set of agents from $1 \dots n$, and let φ be a formula.

$E_G\varphi$ is used to represent that "eveyone in the group G knows φ."

$(M, s) \models E_G\varphi$ if and only if $(M, s) \models K_i\varphi$ for all $i \in G$

$C_G\varphi$ is used to represent that "φ is common knowledge among the agents in G. That is that every member of G knows φ, everyone knows that everyone knows φ, and so on."

Let $E_G^0\varphi$ be an abbreviation for $E_G\varphi$, and $E_G^{k+1}\varphi$ be an abbreviation for $E_G E_G^k\varphi$.

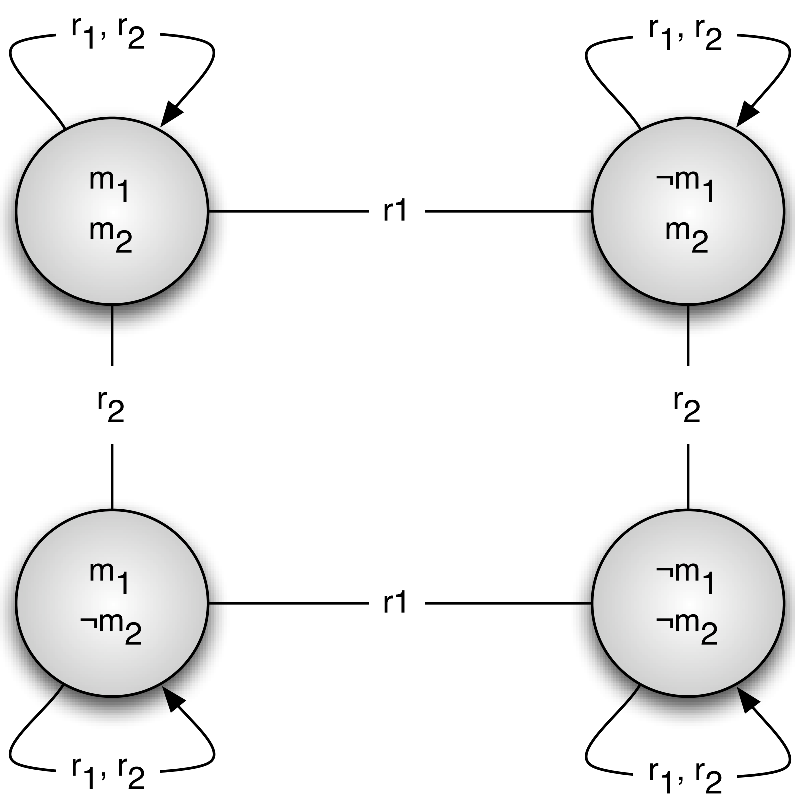
$(M, s) \models C_G\varphi$ if and only if $(M, s) \models E_G^k\varphi$ for $k = 1, 2, 3, \ldots$

The Muddy Children Domain

Suppose that we have 2 children. While playing, both of them get mud on their foreheads. Each child can see the mud on the foreheads of the fellows, but not on themselves. Their father comes by and says "at least one of you has mud on their forehead." He then asks the following question repeatedly: "does any of you know whether you have mud on your own forehead?"

The Domain as a Kripke Structure

We can model the initial state of the agents' knowledge via the following Kripke structure:



Examining this structure we can see that:

- $\triangleright$ $(M, \{m_1, m_2\}) \models K_1 m_2 \wedge K_2 m_1$
- $\triangleright$ $(M, \{m_1, m_2\}) \models \neg K_1 m_1 \wedge \neg K_1 \neg m_1$
- $\triangleright$ $(M, \{\neg m_1, m_2\}) \models \neg K_2 m_2 \wedge \neg K_2 \neg m_2$

Similarly for other states, which corresponds with our intuition and with the problem description.

The Action "Announce"

What happens when the father tells the children that "at least one of them has mud on their foreheads?" We model this by the action:

$$announce(m_1 \vee m_2)$$

Similarly, whenever the children answer no to their father's repeated questions we model this by the action:

$$announce(\neg K_1 m_1 \wedge \neg K_1 \neg m_1 \wedge \neg K_2 m_2 \wedge \neg K_2 \neg m_2)$$

Let φ be a formula. We describe the behavior of the action announce by the following causal law:

$$\mathit{announce}(\varphi) \textbf{ causes } C(\varphi)$$

whose meaning is captured by a transition function from one Kripke structure to another.

Semantics of Announce

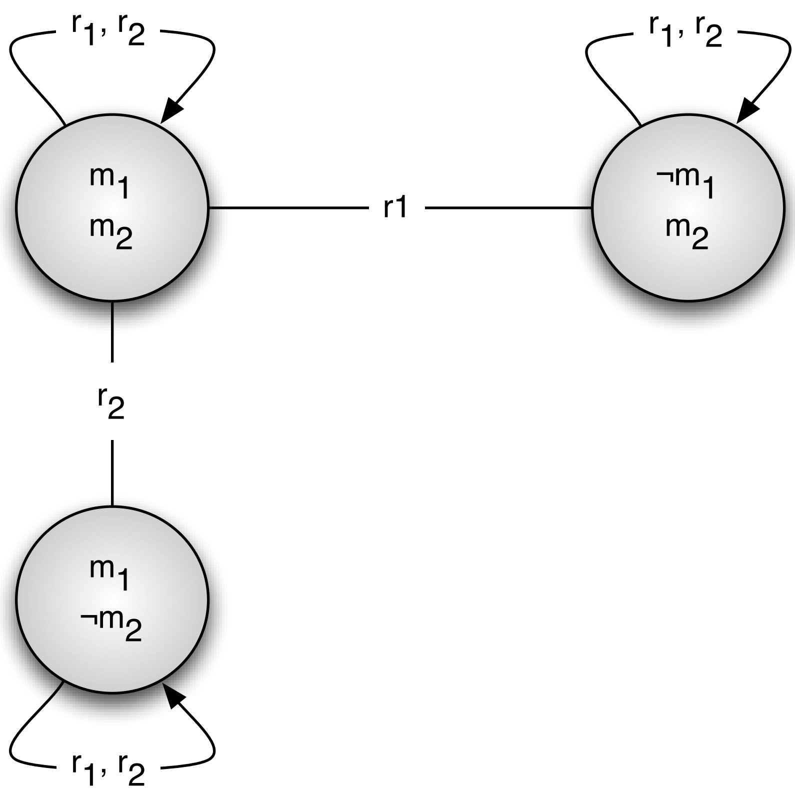
Let φ be a formula, and $M = \langle S, \pi, R_1, \ldots, R_n \rangle$ be our original Kripke structure. We define the successor structure after $announce(\varphi)$ as $M' = \langle S', \pi, R'_1, \ldots, R'_n \rangle$ where:

- ▷ $S' = \{s : s \in S$ and $(M, s) \models \varphi\}$
- ▷ $R'_i = \{(s_1, s_2) : (s_1, s_2) \in R_i$ and $s_1, s_2 \in S'\}$

As an example, after the father announces that at least one of the children is muddy:

$$announce(m_1 \vee m_2)$$

we have the following successor structure:



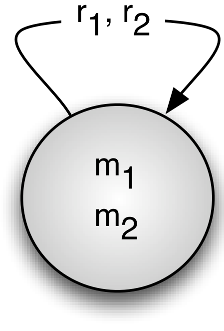
After this first announcement, the father then asks if the children know whether or not they are muddy. This is modeled by the query:

$$ask(\psi) \text{ where } \psi = K_i m_i \vee K_i \neg m_i$$

In order to answer his query, the children use the entailment relation defined above to see if $M \models \psi$. As $M \models \neg\psi$, the children answer "no" which is modeled by:

$$announce(\neg K_1 m_1 \wedge \neg K_1 \neg m_1 \wedge \neg K_2 m_2 \wedge \neg K_2 \neg m_2)$$

which gives us the following successor structure:



After the father asks his question the second time, the children answer "yes", which corresponds to $announce(K_1 m_1 \wedge K_2 m_2)$.

Conclusion

We have shown that solving the muddy children problem may be reduced to reasoning using the entailment relation of modal logic and computing the successor states of the action $announce$.

Future Work

What we have described is in essence an algorithm for reasoning with modal logic which centers along the use of the entailment relation and the generation of successor states.

We plan to show the correctness of this algorithm by proving that our transition function accurately captures the meaning of the causal law. To do this we need to prove that:

$$C(\varphi) \text{ is true in } M'$$

In addition we plan on automating this algorithm through the use of logic programs to describe Kripke structures, and to generate the successor structures after various applications of the action $announce$.

Lastly, we plan on exploring similar problem domains which may involve reasoning with regards to the other modal operators, for example domains which involve agents possibly sharing knowledge.