Multi-Agent Systems

Modal Logic for Multi-Agent Systems (Intro)

Albert-Ludwigs-Universität Freiburg

UNI FREIBURG

Bernhard Nebel, Rolf Bergdoll, and Thorsten Engesser

Winter Term 2019/20

Motivation: Applications of Possible World Semantics

Interpretations of propositional variables can be viewed as possible worlds. Relations between possible worlds can be used to express more interesting concepts:

- Temporal concepts like always, next, ... can be modeled as relations between worlds (Prior, 1957).

Motivation: Applications of Possible World Semantics

Interpretations of propositional variables can be viewed as possible worlds. Relations between possible worlds can be used to express more interesting concepts:

- Temporal concepts like always, next, ... can be modeled as relations between worlds (Prior, 1957).
- Execution of computer program can be modeled as transitions between worlds (Pratt, 1976).

Motivation: Applications of Possible World Semantics

Interpretations of propositional variables can be viewed as possible worlds. Relations between possible worlds can be used to express more interesting concepts:

- Temporal concepts like always, next, ... can be modeled as relations between worlds (Prior, 1957).
- Execution of computer program can be modeled as transitions between worlds (Pratt, 1976).
- Knowledge and belief of an agent can be modeled as truth in all worlds the agent considers possible (Hintikka, 1962).

Motivation: Applications of Possible World Semantics

Interpretations of propositional variables can be viewed as possible worlds. Relations between possible worlds can be used to express more interesting concepts:

- Temporal concepts like always, next, ... can be modeled as relations between worlds (Prior, 1957).
- Execution of computer program can be modeled as transitions between worlds (Pratt, 1976).
- Knowledge and belief of an agent can be modeled as truth in all worlds the agent considers possible (Hintikka, 1962).
- Obligations and permissions can be modeled as truth in all (resp. some) ideal worlds (Kanger, 1957; Hintikka 1957).

Motivation: Applications of Possible World Semantics

Interpretations of propositional variables can be viewed as possible worlds. Relations between possible worlds can be used to express more interesting concepts:

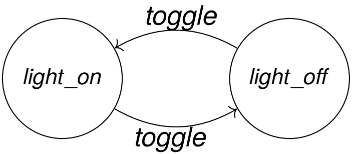
- Temporal concepts like always, next, ... can be modeled as relations between worlds (Prior, 1957).
- Execution of computer program can be modeled as transitions between worlds (Pratt, 1976).
- Knowledge and belief of an agent can be modeled as truth in all worlds the agent considers possible (Hintikka, 1962).
- Obligations and permissions can be modeled as truth in all (resp. some) ideal worlds (Kanger, 1957; Hintikka 1957).
- Desires and intentions can be modeled as truth in all worlds an agent prefers (Cohen & Levesque, 1990).

Kripke Models: Informally

Kripke Model

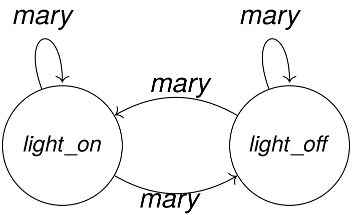
A Kripke model can be viewed as a graph where the nodes represent worlds and the edges represent accessibility between worlds.

Kripke Models: Examples (Programs)



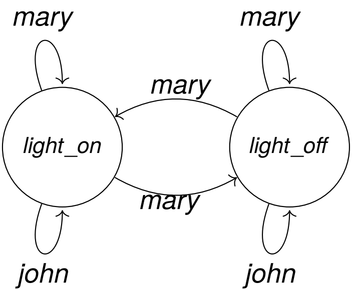
- If the light is on then it is true that after toggling the light is off. If the light is off then it is true that after toggling the light is on.

Kripke Model: Examples (Single-Agent Knowledge)



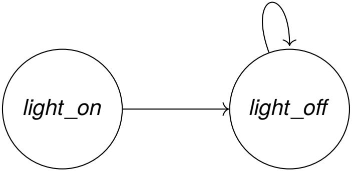
- If the light is on then it is true that Mary considers possible both that the light is on or off. If the light is off then it is true that Mary considers possible both that the light is on or off.

Kripke Model: Examples (Multi-Agent Knowledge



- If the light is on it is true that John only considers possible that the light is on. If the light is off it is true that John only considers possible that the light is off.
- In either world it is true that Mary is uncertain about the state of the switch and John knows about the state of the switch.

Kripke Model: Examples (Permissions)



- If the light is on it is true that it is permissible to bring about that the light is off and it is not permissible to leave the light on.
- If the light is off it is true that it is permissible leave the light off and it is not permissible to bring about that the light is on.
- ⇒In both worlds it is obligatory to bring about/maintain that the light is off.

Kripke Models: Formally

UNI FREIBURG

Kripke Frame

Given a countable set of edge labels $\mathscr{I}$, a Kripke Frame is a tuple (W, R) such that:

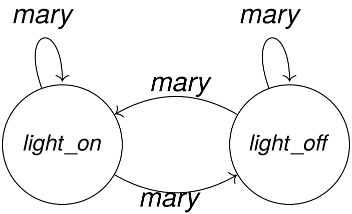
- W is a non-empty set of possible worlds, and
- $R : I \to 2^{W \times W}$ maps each $I \in \mathscr{I}$ to a binary relation $R(I)$ on W (called the accessibility relation of I).

Kripke Model

$M = (W, R, V)$ is a Kripke Model where:

- (W, R) is a Kripke frame, and
- $V : \mathscr{P} \to 2^W$ is called the valuation of a set of node labels $\mathscr{P}$.

Kripke Model: Example



- Kripke Frame (W, R)
  - Possible worlds $W = \{w_l, w_r\}$
  - Edge labels $\mathscr{I} = \{mary\}$
  - $R(mary) = \{(w_l, w_l), (w_l, w_r), (w_r, w_r), (w_r, w_l)\}$
- Kripke Model (W, R, V)
  - W, R as before.
  - Node labels $\mathscr{P} = \{light\_on, light\_off\}$
  - $V(light\_on) = \{w_l\}$, $V(light\_off) = \{w_r\}$

Classes of Kripke Models

- Besides being able to model concrete situations, we are interested in the study of the general properties of concepts like knowledge, intention, obligation etc.
- $\Rightarrow$Identify particular classes of Kripke models as representations of the concept under consideration.
  - Classes of Kripke models can be distinguished based on the properties of their respective frames.
  - **K**: All Kripke frames
  - **T**: Kripke frames with reflexive accessibility relation
  - **D**: Kripke frames with serial accessibility relation
  - **4**: Kripke frames with transitive accessibility relation
  - **5**: Kripke frames with Euclidean accessibility relation
  - Can be combined:
    - **K**, **KD**, **K4**, **K5**, **KT** = **KDT**, **K45**, **KD5**, **KD4**, **KT4** = **KDT4**, **KD45**, **KT5** = **KT45** = **KDT5** = **KDT45**
    - Some abbreviations often used: **KT** is called **T**, **KT4** is called **S4**, **KD45** is weak-S5, **KT5** called **S5**.

Next time: Languages for Talking about Kripke Models

- Kripke models can be described and reasoned about using modal logics.
  - Does a given Kripke model satisfy some given property?
    - E.g., is it currently true that Mary does not know whether the light is on?
  - Do all Kripke models of a class satisfying property A also satisfy property B?
    - E.g., is it always true that if some agent X knows that some agent Y knows Z that agent X knows Z, too?
  - $\Rightarrow$We will learn how to check formulae against given Kripke models, and how to automatically build Kripke models to (dis-)prove a formula's (un-)satisfiability.

Literature

- M. Wooldridge, **An Introduction to MultiAgent Systems**, 2nd Edition, John Wiley & Sons, 2009.
- Y. Shoham, K. Layton-Brown, **Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations**, Cambridge University Press, 2009.
- O. Gasquet, A. Herzig, B. Said, F. Schwarzentruber, **Kripke's Worlds — An Introduction to Modal Logics via Tableaux**, Springer, ISBN 978-3-7643-8503-3, 2014.