Multi-Agent Systems

Albert-Ludwigs-Universität Freiburg

Bernhard Nebel, Felix Lindner, and Thorsten Engesser

Winter Term 2018/19

Overview

- Recap
  - Situations from various domains (Programs, Knowledge, Belief, Desire, Obligation) can nicely be modeled using graphical models.
  - Kripke models formalize graphical models.
  - By constraining the accessibility relations of Kripke frames we obtain classes that correspond to above concepts (Knowledge, Belief etc.)
- Today
  - Introducing formal languages to talk about Kripke models and thus generally about Knowledge, Belief, Desire, Obligation ...

Modal Logics

Kripke Models

UNI FREIBURG

Kripke Frame

Given a countable set of edge labels $\mathcal{I}$, 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 $I \in \mathcal{I}$ 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
- $V : \mathcal{P} \to 2^{W}$ is called the valuation of a set of node labels $\mathcal{P}$.

Formulas of Modal Logic

The language $\mathcal{F}$ of modal logic is inductively defined as follows:

- $\mathcal{P} \subseteq \mathcal{F}$.
- $\{\top, \bot\} \subseteq \mathcal{F}$.
- If $\varphi \in \mathcal{F}$, then $\neg\varphi \in \mathcal{F}$.
- If $\varphi, \psi \in \mathcal{F}$ then $(\varphi \wedge \psi), (\varphi \vee \psi), (\varphi \rightarrow \psi), (\varphi \leftrightarrow \psi) \in \mathcal{F}$
- If $\varphi \in \mathcal{F}$ and $I \in \mathcal{I}$, then $\mathbf{[I]}\varphi, \mathbf{\langle I \rangle}\varphi \in \mathcal{F}$.

Different Variants of Languages

- Alethic logic (Necessity): $\Box, \Diamond$
- Epistemic logic (Knowledge): $\mathbf{K}$, $\hat{\mathbf{K}}$
- Doxastic logic (Belief): $\mathbf{B}$, $\hat{\mathbf{B}}$
- Deontic logic (Obligation): $\mathbf{O}$, $\mathbf{P}$
- Multi-Agent Epistemic logic: Agent name as subscript, e.g., $\mathbf{K}_{mary}\hat{\mathbf{K}}_{john} sun\_shining$

- Notation: Sometimes, we will decide that **[I]** shall be read in context of epistemic logic, sometimes we will decide to read it in context of deontic logic. We then may also sometimes write K_I (i.e., Agent I knows) , and O_I (i.e., Agent I ought to) instead of **[I]**.

Semantics: Truth Conditions

Given a Kripke model M, a possible world w of M, and a formula φ. We define when φ is true at w, written $M, w \models \varphi$:

- $M, w \models p$ iff. $w \in V(p)$, for atomic formulae $p \in \mathcal{P}$.
- $M, w \not\models \bot$.
- $M, w \models \top$.
- $M, w \models \neg\varphi$ iff. $M, w \not\models \varphi$.
- $M, w \models (\varphi \wedge \psi)$ iff. $M, w \models \varphi$ and $M, w \models \psi$.
- $M, w \models (\varphi \vee \psi)$ iff. $M, w \models \varphi$ or $M, w \models \psi$.
- $M, w \models (\varphi \rightarrow \psi)$ iff. $M, w \not\models \varphi$ or $M, w \models \psi$.
- $M, w \models (\varphi \leftrightarrow \psi)$ iff. $M, w \models (\varphi \rightarrow \psi)$ and $M, w \models (\psi \rightarrow \varphi)$.
- $M, w \models \mathbf{[I]}\varphi$ iff. for every u: if $(w, u) \in R(I)$ then $M, u \models \varphi$.
- $M, w \models \mathbf{\langle I \rangle}\varphi$ iff. for some u: $(w, u) \in R(I)$ and $M, u \models \varphi$.

Duality

1. $M, w \models [\mathbf{I}]\varphi$ iff. $M, w \models \neg\langle\mathbf{I}\rangle\neg\varphi$
2. $M, w \models \langle\mathbf{I}\rangle\varphi$ iff. $M, w \models \neg[\mathbf{I}]\neg\varphi$

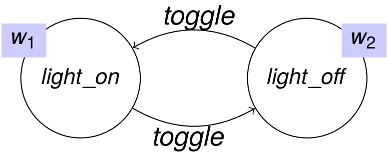
- To see that (1):
  - $M, w \models [\mathbf{I}]\varphi$
  - iff. for every u: if $(w, u) \in R(I)$ then $M, u \models \varphi$
  - iff. it is not the case that for some u: $(w, u) \in R(I)$ and $M, u \models \neg\varphi$
  - iff. not $M, w \models \langle\mathbf{I}\rangle\neg\varphi$
  - iff. $M, w \models \neg\langle\mathbf{I}\rangle\neg\varphi$

Model Checking

- Question: Is a given formula φ true in world w in model M?
- Input: A Kripke model M, a world w in M, and a formula φ.
- Output: “Yes” if $M, w \models \varphi$, “No” else.

Model Checking: Example

M:



- $M, w_1 \models \langle toggle \rangle \top \wedge [toggle][toggle] light\_on$

1 $M, w_1 \models \langle toggle \rangle \top$

1.1 for some u: $(w_1, u) \in R(toggle)$ and $M, u \models \top$.

1.1.1 we find $(w_1, w_2) \in R(toggle)$ and $M, w_2 \models \top$.☺

2 $M, w_1 \models [toggle][toggle] light\_on$

2.1 for every u: if $(w_1, u) \in R(toggle)$ then $M, u \models [toggle] light\_on$

2.1.1 $M, w_2 \models [toggle] light\_on$.

2.1.1.1 for every u: if $(w_2, u) \in R(toggle)$ then $M, u \models light\_on$

2.1.1.1.1 $M, w_1 \models light\_on$. ☺

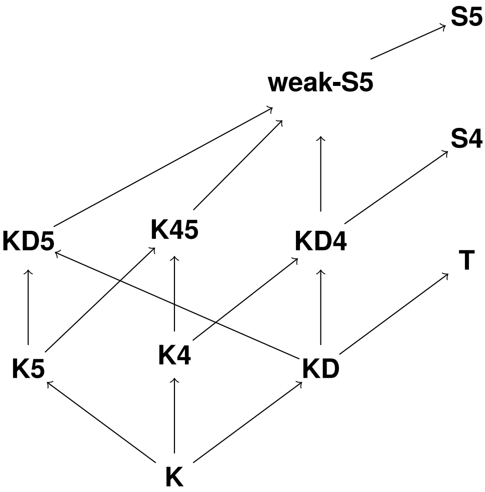
Semantics: Satisfiability

- A formula φ is satisfiable in a model $M = (W, R, V)$, if there exists a world $w \in W$ such that $M, w \models \varphi$.
- A formula φ is satisfiable in a frame if it is satisfiable in a model based on that frame.
- A formula φ is satisfiable in a class of frames if it is satisfiable in a model based on some frame from the class of frames.
- A formula φ is true in a model M ($M \models \varphi$) if φ is true in all worlds of M.

Semantics: Validity

- A formula is valid in a frame if φ is true in all models based on that frame.
- We say that a formula φ is valid in a class of frames **C** (**K**, **T**, **D**, **4**, **5**, and combinations thereof), written $\models_{\mathbf{C}} \varphi$, iff. $(W,R,V),w \models \varphi$
  - for every frame (W,R),
  - every valuation V over (W,R),
  - every world w in W.

A Lattice of Classes



Validity in a Class of Frames

- Valid formulas give us an idea of how the classes differ, and thus what is and is not specific to the general behavior of our modalities (Knowledge, Belief, Obligation etc.).

- Correspondences between classes of frames and formulas
  - **[I]**$(\varphi \to \psi) \to ($**[I]**$\varphi \to$ **[I]**$\psi)$ (for every formulae φ, ψ) is **K**-valid (valid in the class of all frames)
  - **[I]**$\varphi \to \varphi$ (for every formulae φ) is **T**-valid (exactly valid in the class of reflexive frames)
  - **[I]**$\varphi \to$ **<I>**φ (for every formulae φ) is **D**-valid (exactly valid in the class of serial frames)
  - **[I]**$\varphi \to$ **[I][I]**φ (for every formulae φ) is **4**-valid (exactly valid in the class of transitive frames)
  - **<I>**$\varphi \to$ **[I]<I>**φ (for every formulae φ) is **5**-valid (exactly valid in the class of Euclidean frames)

Validity in a Class of Frames: Example I

- $\textbf{[I]}(\varphi \rightarrow \psi) \rightarrow (\textbf{[I]}\varphi \rightarrow \textbf{[I]}\psi)$ is **K**-valid:
  1 Let M be a arbitrarily chosen Kripke model and w be a arbitrary world in M.
  1.1 Assume $M,w \models \textbf{[I]}(\varphi \rightarrow \psi)$ (otherwise the formula is true anyway ☺). Thus, for every world u: if $(w,u) \in R(I)$ then $M,u \models \varphi \rightarrow \psi$.
  1.1.1 If $\textbf{[I]}\varphi$ is false in w, then $(\textbf{[I]}\varphi \rightarrow \textbf{[I]}\psi)$ is true in w, and the overall formula is true in w. ☺
  1.1.2 If $\textbf{[I]}\varphi$ is true in w, then both $\textbf{[I]}\varphi \rightarrow \psi$ and $\textbf{[I]}(\varphi)$ are true in w. Thus, in every world u accessible from w, also ψ is true, i.e., $\textbf{[I]}(\psi)$ is true in w. Therefore, the overall formula is true in w. ☺

Validity in a Class of Frames: Example II

- $[\mathbf{I}]\varphi \rightarrow \varphi$ is not **K**-valid:
  - Consider Kripke model $M = (W, R, V)$ from class **K**:
    - $W = \{w\}$
    - $R(I) = \{\}$
    - $V(p) = \{\}$
  - Check that $M, w \models [\mathbf{I}]p$ and $M, w \not\models p$. Thus, $M, w \not\models [\mathbf{I}]p \rightarrow p$.

Entailment

- A formula φ entails ψ in the class $\mathbf{C}$ (written $\varphi \models_{\mathbf{C}} \psi$) iff for every model M based on some frame in $\mathbf{C}$ and every possible world w of M:
  - if $M,w \models_{\mathbf{C}} \varphi$ then $M,w \models_{\mathbf{C}} \psi$
- Entailment can be reduced to validity:
  - $\varphi \models_{\mathbf{C}} \psi$ iff. $\models_{\mathbf{C}} \varphi \rightarrow \psi$
  - $\Theta \models_{\mathbf{C}} \psi$ iff. $\models_{\mathbf{C}} \bigwedge \Theta \rightarrow \psi$

Reducing Validity to Satisfiability

- The validity problem can be reduced to the satisfiability problem:
  - Instead of asking whether φ is true in all worlds in all Kripke models in a class, we can ask if $\neg\varphi$ is true in some world in some Kripke model in the class.
- Problem formulation:
  - Input: A formula φ.
  - Output: "Yes" if there is a Kripke model M and a world w of M such that $M, w \models \varphi$, "No" otherwise.
- It turns out that we can systematically search for Kripke models that satisfy some formula. With this tool at hand, we can algorithmically decide validity.

Literature I

M. Wooldridge, An Introduction to MultiAgent Systems, John Wiley & Sons, 2002.

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.