Principles of Knowledge Representation and Reasoning

Modal Logics

Bernhard Nebel, Felix Lindner, and Thorsten Engesser

November 9 & 11, 2015

1 Motivation

Motivation for studying modal logics

- Notions like believing and knowing require a more general semantics than e.g. propositional logic has.
- Some KR formalisms can be understood as (fragments of) a propositional modal logic.
- Application 1: Spatial representation formalism RCC8
- Application 2: Description logics
- Application 3: Reasoning about time
- Application 4: Reasoning about actions, strategies, etc.

Motivation for modal logics

Often, we want to state something where we have an "embedded proposition":

- John believes that it is Sunday.
- I know that $2^{10} = 1024$.

Reasoning with embedded propositions:

- *John believes that if it is Sunday, then shops are closed.*
- *John believes that it is Sunday.*
- This implies (assuming belief is closed under modus ponens):
  John believes that shops are closed.

$\rightsquigarrow$ How to formalize this?

UNI FREIBURG

2 Syntax

Syntax

Propositional logic $+$ operators $\Box$ & $\Diamond$ (Box & Diamond):

$$\begin{array}{llll} \varphi & ::= & \ldots & \text{classical propositional formula} \\ & | & \Box\varphi' & \text{Box} \\ & | & \Diamond\varphi' & \text{Diamond} \end{array}$$

$\Box$ and $\Diamond$ have the same operator precedence as $\neg$.

Some possible readings of $\Box\varphi$:

- Necessarily φ (alethic)
- Always φ (temporal)
- φ should be true (deontic)
- Agent A believes that φ (doxastic)
- Agent A knows that φ (epistemic)

$\rightsquigarrow$ Different semantics for different intended readings

3 Semantics

- Possible worlds
- Kripke semantics
- Basic notions
- Relational properties vs. axioms

Truth-functional semantics?

- Is it possible to define the meaning of $\Box\varphi$ truth-functionally, i.e. by referring to the truth value of φ only?
- An attempt to interpret necessity truth-functionally:
  - If φ is false, then $\Box\varphi$ should be false.
  - If φ is true, then …
    - …$\Box\varphi$ should be true $\rightsquigarrow$ $\Box$ is the identity function
    - …$\Box\varphi$ should be false $\rightsquigarrow$ $\Box\varphi$ is identical to falsity
- Note: There are only 4 different unary Boolean functions $\{T,F\} \to \{T,F\}$.

UNI FREIBURG

Semantics: the idea

In classical propositional logic, formulae are interpreted over single interpretations and are evaluated to true or false.

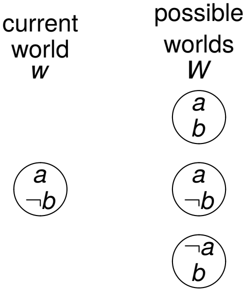
In modal logics one considers sets of interpretations: possible worlds (physically possible, conceivable, ...).

Main idea:

- Consider a world (interpretation) w and a set of worlds W which are possible with respect to w.
- A classical formula (with no modal operators) φ is true with respect to (w, W) iff φ is true in w.
- $\Box\varphi$ is true wrt. (w, W) iff φ is true in all worlds in W.
- $\Diamond\varphi$ is true wrt. (w, W) iff φ is true in some world in W.
- Meanings of $\Box$ and $\Diamond$ are interrelated by: $\Diamond\varphi \equiv \neg\Box\neg\varphi$.

UNI FREIBURG

Semantics: an example



Examples:

- $a \wedge \neg b$ is true relative to (w, W).
- $\Box a$ is not true relative to (w, W).
- $\Box(a \vee b)$ is true relative to (w, W).

Question: How to evaluate modal formulae in $w \in W$?

$\rightsquigarrow$ For each world, we specify a set of possible worlds.

$\rightsquigarrow$ Frames

Frames, interpretations, and worlds

Definition (Kripke frame)

A (Kripke, relational) frame is a pair $\mathcal{F} = \langle W, R \rangle$, where W is a non-empty set (of worlds) and $R \subseteq W \times W$ is a binary relation on W (accessibility relation).

For $(w, v) \in R$ we write also $w\,R\,v$. We say that v is an R-successor of w or that v is R-reachable from w.

Definition (Kripke model)

For a given set of propositional variables Σ, a Kripke model (or interpretation) based on the frame $\mathcal{F} = \langle W, R \rangle$ is a triple $\mathcal{I} = \langle W, R, \pi \rangle$, where π is a function that maps worlds w to truth assignments $\pi_w : \Sigma \to \{T, F\}$, i.e.:

$$\pi : W \to \{T, F\}^{\Sigma}, \;\; w \mapsto \pi_w.$$

Semantics: truth in a world

A formula φ is true in world w in an interpretation $\mathcal{I} = \langle W, R, \pi \rangle$ under the following conditions:

$$
\begin{array}{lll}
\mathcal{I}, w \models a & \text{iff} & \pi_w(a) = T \\
\mathcal{I}, w \models \top & & \\
\mathcal{I}, w \not\models \bot & & \\
\mathcal{I}, w \models \neg\varphi & \text{iff} & \mathcal{I}, w \not\models \varphi \\
\mathcal{I}, w \models \varphi \wedge \psi & \text{iff} & \mathcal{I}, w \models \varphi \text{ and } \mathcal{I}, w \models \psi \\
\mathcal{I}, w \models \varphi \vee \psi & \text{iff} & \mathcal{I}, w \models \varphi \text{ or } \mathcal{I}, w \models \psi \\
\mathcal{I}, w \models \varphi \rightarrow \psi & \text{iff} & \mathcal{I}, w \not\models \varphi \text{ or } \mathcal{I}, w \models \psi \\
\mathcal{I}, w \models \varphi \leftrightarrow \psi & \text{iff} & \mathcal{I}, w \models \varphi \text{ if and only if } \mathcal{I}, w \models \psi \\
\mathcal{I}, w \models \Box\varphi & \text{iff} & \mathcal{I}, u \models \varphi, \text{ for all } u \text{ s.t. } wRu \\
\mathcal{I}, w \models \Diamond\varphi & \text{iff} & \mathcal{I}, u \models \varphi, \text{ for some } u \text{ s.t. } wRu
\end{array}
$$

Satisfiability and validity

A formula φ is satisfiable in an interpretation $\mathcal{I}$ if there exists a world w in $\mathcal{I}$ such that $\mathcal{I}, w \models \varphi$.

A formula φ is satisfiable in a frame $\mathcal{F}$ (satisfiable in a class of frames $\mathcal{C}$) if it is satisfiable in an interpretation $\mathcal{I}$ based on $\mathcal{F}$ (satisfiable in an interpretation $\mathcal{I}$ based on some frame contained in $\mathcal{C}$).

A formula φ is true in an interpretation $\mathcal{I}$ (symbolically $\mathcal{I} \models \varphi$) if φ is true in all worlds of $\mathcal{I}$.

A formula φ is valid in a frame $\mathcal{F}$ or $\mathcal{F}$-valid (symb. $\mathcal{F} \models \varphi$) if φ is true in all interpretations based on $\mathcal{F}$.

A formula φ is valid in a class of frames $\mathcal{C}$ or $\mathcal{C}$-valid (symb. $\mathcal{C} \models \varphi$) if $\mathcal{F} \models \varphi$ for all $\mathcal{F} \in \mathcal{C}$.

Validities in $\mathbf{K}$

K denotes the class of all frames – named after Saul Kripke, who invented this semantics.

Some validities in **K**:

1. $\varphi \vee \neg\varphi$
2. $\Box(\varphi \vee \neg\varphi)$
3. $\Box\varphi$, if φ is a classical tautology
4. $\Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$ (axiom schema *K*)

Moreover, it holds:

If φ is **K**-valid, then $\Box\varphi$ is **K**-valid

Validity: some examples

Theorem

*K is **K**-valid.* $K = \Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$

Proof.

Let $\mathcal{I}$ be an interpretation and let w be a world in $\mathcal{I}$.
Assume $\mathcal{I}, w \models \Box(\varphi \rightarrow \psi)$, i.e., in all worlds u with wRu, if φ is true then also ψ is. (Otherwise K is true in w anyway.)
If $\Box\varphi$ is false in w, then $(\Box\varphi \rightarrow \Box\psi)$ is true and K is true in w.
If $\Box\varphi$ is true in w, then both $\Box(\varphi \rightarrow \psi)$ and $\Box\varphi$ are true in w. Hence both $\varphi \rightarrow \psi$ and φ are true in every world u accessible from w. Hence ψ is true in any such u, and therefore $w \models \Box\psi$.
Since $\mathcal{I}$ and w were chosen arbitrarily, the argument goes through for any $\mathcal{I}, w$, i.e., K is **K**-valid. $\Box$

Non-validity: example

Proposition

*$\Diamond\top$ is not **K**-valid.*

Proof.

A counterexample is the following interpretation $\mathcal{I} = \langle W, R, \pi \rangle$ with:

$$W := \{w\},$$
$$R := \emptyset,$$
$$\pi_w(a) := T \quad (a \in \Sigma).$$

We have $\mathcal{I}, w \not\models \Diamond\top$ because there is no u such that wRu. $\square$

UNI FREIBURG

Non-validity: example

Proposition

$\Box\varphi \rightarrow \varphi$ is not **K***-valid.*

Proof.

A counterexample is the following interpretation $\mathcal{I} = \langle W, R, \pi \rangle$ with:

$$\begin{aligned} W &:= \{w\}, \\ R &:= \emptyset, \\ \pi_w(a) &:= F \quad (a \in \Sigma). \end{aligned}$$

We have $\mathcal{I}, w \models \Box a$, but $\mathcal{I}, w \not\models a$. □

Non-validity: another example

Proposition

*$\Box\varphi \rightarrow \Box\Box\varphi$ is not **K**-valid.*

Proof.

A counterexample is the following interpretation:

$$\mathcal{I} = \langle\{u,v,w\},\{(u,v),(v,w)\},\pi\rangle$$

with

$$\pi_u(a) := T$$
$$\pi_v(a) := T$$
$$\pi_w(a) := F$$

Hence, $\mathcal{I},u \models \Box a$, but $\mathcal{I},u \not\models \Box\Box a$. $\Box$

Accessibility and axiom schemata

Let us consider the following axiom schemata:

T:	$\Box\varphi \rightarrow \varphi$	(knowledge axiom)
4:	$\Box\varphi \rightarrow \Box\Box\varphi$	(positive introspection)
5:	$\Diamond\varphi \rightarrow \Box\Diamond\varphi$	(or $\neg\Box\varphi \rightarrow \Box\neg\Box\varphi$: negative introspection)
B:	$\varphi \rightarrow \Box\Diamond\varphi$	
D:	$\Box\varphi \rightarrow \Diamond\varphi$	(or $\Box\varphi \rightarrow \neg\Box\neg\varphi$: disbelief in the negation)

… and the following classes of frames, for which the accessibility relation is restricted as follows:

- T: reflexive (*wRw* for each world *w*)
- 4: transitive (*wRu* and *uRv* implies *wRv*)
- 5: euclidian (*wRu* and *wRv* implies *uRv*)
- B: symmetric (*wRu* implies *uRw*)
- D: serial (for each *w* there exists *v* with *wRv*)

Correspondence between accessibility relations and axiom schemata (1)

Theorem

Axiom schema T (4, 5, B, D) is **T**- *valid (***4-, 5-, B-***, or* **D***-valid, respectively).*

Proof.

For T and **T**: Let $\mathcal{F}$ be a frame from class **T**. Let $\mathcal{I}$ be an interpretation based on $\mathcal{F}$ and let w be an arbitrary world in $\mathcal{I}$.
If $\Box\varphi$ is not true in world w, then axiom T is true in w.
If $\Box\varphi$ is true in w, then φ is true in all accessible worlds. Since the accessibility relation is reflexive, w is among the accessible worlds, i.e., φ is true in w. Thus also in this case T is true in w.
We conclude: T is true in all worlds in all interpretations based on **T**-frames. □

UNI FREIBURG

Correspondence between accessibility relations and axiom schemata (2)

Theorem

If T $(4, 5, B, D)$ is valid in a frame $\mathcal{F}$, then $\mathcal{F}$ is a **T***-frame (***4-, 5-, B-***, or* **D***-frame, respectively).*

Proof.

For T and **T**: : Assume that $\mathcal{F}$ is not a **T**-frame. We will construct an interpretation based on $\mathcal{F}$ that falsifies T.
Because $\mathcal{F}$ is not a **T**-frame, there is a world w such that not wRw.
Construct an interpretation $\mathcal{I}$ such that $\mathcal{I}, w \not\models a$ and $\mathcal{I}, v \models a$ for all v such that wRv.
Now $\mathcal{I}, w \models \Box a$ and $\mathcal{I}, w \not\models a$, and hence $\mathcal{I}, w \not\models \Box a \rightarrow a$. $\square$

UNI FREIBURG

4 Different Logics

Different modal logics

Name	Property	Axiom schema
K	–	$\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$
T	reflexivity	$\Box\varphi \to \varphi$
4	transitivity	$\Box\varphi \to \Box\Box\varphi$
5	euclidicity	$\Diamond\varphi \to \Box\Diamond\varphi$
B	symmetry	$\varphi \to \Box\Diamond\varphi$
D	seriality	$\Box\varphi \to \Diamond\varphi$

Some basic modal logics:

$$
\begin{array}{rcl}
K & & \\
KT4 & = & S4 \\
KT5 & = & S5 \\
\vdots & &
\end{array}
$$

Different modal logics

logics	$\Box$	$\Diamond = \neg\Box\neg$	K	T	4	5	B	D
alethic	necessarily	possibly	Y	Y	Y	Y	Y	Y
epistemic	known	possible	Y	Y	Y	Y	Y	Y
doxastic	believed	possible	Y	N	Y	Y	N	Y
deontic	obligatory	permitted	Y	N	Y?	Y?	N	Y
temporal	always (in the future)	sometimes (...)	Y	Y/N	Y	N	N	N/Y

5 Analytic Tableaux

- Tableau rules

Proof methods

- How can we show that a formula is $\mathcal{C}$-valid?
- In order to show that a formula is not $\mathcal{C}$-valid, one can construct a counterexample (= an interpretation that falsifies it).
- When trying out all ways of generating a counterexample without success, this counts as a proof of validity.

$\rightsquigarrow$ Method of (analytic/semantic) tableaux

UNI FREIBURG

Tableaux method

A tableau is a tree with nodes marked as follows:

- $w \models \varphi$,
- $w \not\models \varphi$, and
- wRv.

A branch that contains nodes marked with $w \models \varphi$ and $w \not\models \varphi$ is closed. All other branches are open. If all branches are closed, the tableau is called closed.

A tableau is constructed by using the tableau rules.

Tableau rules for propositional logic

$$\frac{w \models \varphi \vee \psi}{w \models \varphi \mid w \models \psi} \qquad \frac{w \not\models \varphi \vee \psi}{\begin{array}{c} w \not\models \varphi \\ w \not\models \psi \end{array}} \qquad \frac{w \models \neg\varphi}{w \not\models \varphi}$$

$$\frac{w \models \varphi \wedge \psi}{\begin{array}{c} w \models \varphi \\ w \models \psi \end{array}} \qquad \frac{w \not\models \varphi \wedge \psi}{w \not\models \varphi \mid w \not\models \psi} \qquad \frac{w \not\models \neg\varphi}{w \models \varphi}$$

$$\frac{w \models \varphi \rightarrow \psi}{w \not\models \varphi \mid w \models \psi} \qquad \frac{w \not\models \varphi \rightarrow \psi}{\begin{array}{c} w \models \varphi \\ w \not\models \psi \end{array}}$$

Additional tableau rules for modal logic **K**

$$\frac{w \models \Box\varphi}{v \models \varphi}$$ if wRv is on the branch already

$$\frac{w \not\models \Box\varphi}{\begin{array}{c} wRv \\ v \not\models \varphi \end{array}}$$ for new v

$$\frac{w \models \Diamond\varphi}{\begin{array}{c} wRv \\ v \models \varphi \end{array}}$$ for new v

$$\frac{w \not\models \Diamond\varphi}{v \not\models \varphi}$$ if wRv is on the branch already

Properties of **K** tableaux

Proposition

If a K-tableau is closed, the truth condition at the root cannot be satisfied.

Theorem (Soundness)

If a K-tableau with root $w \not\models \varphi$ is closed, then φ is **K***-valid.*

Theorem (Completeness)

If φ is **K***-valid, then there is a closed tableau with root $w \not\models \varphi$.*

Termination: There are strategies for constructing **K**-tableaux that always terminate after a finite number of steps, and result in a closed tableau whenever one exists.

UNI FREIBURG

Tableau rules for other modal logics

Proofs within more restricted classes of frames allow the use of further tableau rules.

- For reflexive (**T**) frames we may extend any branch with wRw.
- For transitive (**4**) frames we have the following additional rule:
  - If wRv and vRu are in a branch, wRu may be added to the branch.
- For serial (**D**) frames we have the following rule:
  - If there is $w \models \ldots$ or $w \not\models \ldots$ on a branch, then add wRv for a new world v.
- Similar rules for other properties...

Complextity of simple modal logics

How hard is it to check whether a modal logic formula is satisfiable or valid?

The answer depends in fact on the considered class of frames!

For example, one can show that each formula φ that is satisfiable in some S5-frame is satisfiable in an S5-frame with $|W| \leq |\varphi|$.

Proposition

Checking whether a modal formula is satisfiable in some S5-model is NP-complete (and hence checking S5-validity is coNP-complete).

For other modal logics, such as K, KT, KD, K4, S4, these problems are PSPACE-complete.

6 Embedding in FOL

Connection between propositional modal logic and FOL?

- There are similarities between predicate logic and propositional modal logics:
  1. $\Box$ vs. $\forall$
  2. $\Diamond$ vs. $\exists$
  3. possible worlds vs. objects of the universe
- In fact, many propositional modal logics can be embedded in the predicate logic.

$\Rightarrow$ Modal logics can be understood as a sublanguage of FOL.

UNI FREIBURG

Embedding modal logics into FOL (1)

1. $\tau(p,x) = p(x)$ for propositional variables p
2. $\tau(\neg\varphi,x) = \neg\tau(\varphi,x)$
3. $\tau(\varphi \vee \psi,x) = \tau(\varphi,x) \vee \tau(\psi,x)$
4. $\tau(\varphi \wedge \psi,x) = \tau(\varphi,x) \wedge \tau(\psi,x)$
5. $\tau(\Box\varphi,x) = \forall y(R(x,y) \rightarrow \tau(\varphi,y))$ for some new y
6. $\tau(\Diamond\varphi,x) = \exists y(R(x,y) \wedge \tau(\varphi,y))$ for some new y

Embedding modal logics into FOL (2)

Theorem

φ is K-valid if and only if $\forall x\, \tau(\varphi, x)$ is valid in FOL.

Theorem

φ is T-valid if and only if in FOL the logical consequence $\{\forall x R(x,x)\} \models \forall x\, \tau(\varphi, x)$ holds.

Example

$\Box p \wedge \Diamond(p \rightarrow q) \rightarrow \Diamond q$ is K-valid, because

$$\forall x(\forall x'(R(x,x') \rightarrow p(x')) \wedge \exists x'(R(x,x') \wedge (p(x') \rightarrow q(x'))) \\ \rightarrow \exists x'(R(x,x') \wedge q(x')))$$

is valid in FOL.

7 Outlook & literature

UNI FREIBURG

Outlook

We only looked at some basic propositional modal logics. There are also:

- modal first order logics (with quantification $\forall$ and $\exists$ and predicates)
- multi-modal logics: more than one modality, e.g. knowledge/belief operators for several agents
- temporal and dynamic logics (modalities that refer to time or programs, respectively)

Outlook

Did we really do something new? Couldn't we have done everything in propositional modal logic in FOL already?

- Yes – but now we know much more about the (restricted) system and have decidable problems!

Literature I

Anil Nerode.
Some lectures on modal logic.
In F. L. Bauer, editor, **Logic, Algebra, and Computation**, NATO ASI Series on Computer and System Sciences, pages 281–334. Springer, 1991.

Melvin Fitting.
Basic Modal Logic.
In D. M. Gabbay and C. J. Hogger and J. A. Robinson, eds., **Handbook of Logic in Artificial Intelligence and Logic rogramming – Vol. 1: Logical Foundations**, Oxford University Press, 1993.

P. Blackburn, P., M. de Rijke, and Y. Venema.
Modal Logic.
Cambridge Tracks in Theoretical Computer Science. Cambridge University Press, 2002.

Literature II

M. Fitting.
Proof Methods for Modal and Intuitionistic Logic.
Reidel, 1983.

Robert Goldblatt.
Logics of Time and Computation.
Stanford University, 1992.

B. F. Chellas.
Modal Logic: An Introduction.
Cambridge University, 1980.

J. Y. Halpern, R. Fagin, Y. Moses, and M. Y. Vardi
Reasoning About Knowledge.
MIT Press, 1995.