FUNDAMENTALS OF ARTIFICIAL INTELLIGENCE

Exercise 4: Logical Agents – Problems   Winter semester 2023/2024

Florian Lercher

Problem 4.1: Model, satisfaction relation, and entailment

(*Taken from [1] question 7.4*) Which of the following statements are correct? Prove correctness by reasoning about the models satisfying each sentence.

1. $\mathit{False} \models \mathit{True}$
2. $\mathit{True} \models \mathit{False}$
3. $(A \land B) \models (A \Leftrightarrow B)$
4. $(A \Leftrightarrow B) \models (A \lor B)$
5. $(A \Leftrightarrow B) \models (\neg A \lor B)$

Problem 4.2: Validity, satisfiability, and unsatisfiability

(*Exercise is adapted from [1] question 7.10.*) Recall first the definition of *validity* and *satisfiability*.

Problem 4.2.1: Prove the following two metatheorems:

1. Sentence α is valid if and only if $\alpha \equiv \mathit{True}$,
2. Sentence α is unsatisfiable if and only if $\alpha \equiv \mathit{False}$.

Problem 4.2.2: Show whether each of the following sentences is valid, satisfiable, or unsatisfiable. To this end, use the two metatheorems above, the standard logical equivalences from the lecture, and the following four logical equivalences:

$$\alpha \lor \neg\alpha \equiv \mathit{True} \qquad \alpha \lor \alpha \equiv \alpha$$
$$\alpha \land \neg\alpha \equiv \mathit{False} \qquad \alpha \land \alpha \equiv \alpha$$

1. $\mathit{Smoke} \Rightarrow \mathit{Smoke}$
2. $(\mathit{Smoke} \Rightarrow \mathit{Fire}) \Rightarrow (\neg\mathit{Smoke} \Rightarrow \neg\mathit{Fire})$
3. $\mathit{Smoke} \lor \mathit{Fire} \lor \neg\mathit{Fire}$
4. $(\mathit{Fire} \Rightarrow \mathit{Smoke}) \land \mathit{Fire} \land \neg\mathit{Smoke}$

Problem 4.3: Knights and Knaves

(*Proof by inference rule.*) The following puzzle is taken from [2]. Suppose we are on an island with two types of inhabitants: "knights" who always tell the truth, and "knaves" who always lie.

> According to this problem, three of the inhabitants – A, B and C – were standing together in the garden. A stranger passed by and asked A, "Are you a knight or a knave?". A answered, but rather indistinctly, so the stranger could not make out what he said. The stranger then asked B, "What did A say?". B replied, "A said that he is a knave". At this point the third man, C, said "Don't believe B; he's lying!". The question is, what are B and C?

Model this logic puzzle by introducing three atomic propositions A, B, and C with intended interpretation that A, B, and C are knights.

Problem 4.1: Model, satisfaction relation, and entailment

(*Taken from [1] question 7.4*) Which of the following statements are correct? Prove correctness by reasoning about the models satisfying each sentence.

1. $False \models True$
2. $True \models False$
3. $(A \wedge B) \models (A \Leftrightarrow B)$
4. $(A \Leftrightarrow B) \models (A \vee B)$
5. $(A \Leftrightarrow B) \models (\neg A \vee B)$

1. $\equiv M(False) \subseteq M(True)$

$\equiv \phi \subseteq$ All modle

$\equiv$ correct

2. $\equiv M(True) \subseteq M(False)$

$\equiv$ All modle $\subseteq \phi$

$\equiv$ wrong

3

A	B	$A \wedge B$	$A \Leftrightarrow B$
T	T	T	T
T	F	F	F
F	T	F	F
F	F	F	T

$\equiv M(A \wedge B) \subseteq M(A \Leftrightarrow B)$

$\equiv \{(T,T)\} \subseteq \{(T,T), (F,F)\}$

$\equiv$ correct

4

A	B	$A \Leftrightarrow B$	$A \vee B$
T	T	T	T
T	F	F	T
F	T	F	T
F	F	T	F

$\equiv M(A \Leftrightarrow B) \subseteq M(A \vee B)$

$\equiv \{(T,T), (F,F)\} \subseteq \{(T,T), (T,F), (F,T)\}$

$\equiv$ Wrong

5

A	B	$A \Leftrightarrow B$	$\neg A \vee B$
T	T	T	T
T	F	F	F
F	T	F	T
F	F	T	T

$\equiv M(A \Leftrightarrow B) \subseteq M(\neg A \vee B)$

$\equiv \{(T,T), (F,F)\} \subseteq \{(T,T), (F,T), (F,F)\}$

$\equiv$ correct

Problem 4.2.1: Prove the following two metatheorems:

1. Sentence α is valid if and only if $\alpha \equiv True$,
2. Sentence α is unsatisfiable if and only if $\alpha \equiv False$.

1. α is valid i.f.f $\underline{\alpha \equiv True}$

$\equiv \alpha \models True$ and $True \models \alpha$.

$\equiv M(\alpha) \subseteq M(True)$ and $M(True) \subseteq M(\alpha)$

$\equiv M(\alpha) = M(True) =$ All modle

$\equiv \alpha$ is valid

2. α is unsatisfiable i.f.f $\underline{\alpha \equiv False}$

$\equiv \alpha \models False$ and $False \models \alpha$.

$\equiv M(\alpha) \models M(False)$ and $M(False) \models M(\alpha)$

$\equiv M(\alpha) = M(False) = \phi$

$\equiv \alpha$ is unsatisfiable

Problem 4.2.2: Show whether each of the following sentences is valid, satisfiable, or unsatisfiable. To this end, use the two metatheorems above, the standard logical equivalences from the lecture, and the following four logical equivalences:

$$\alpha \vee \neg\alpha \equiv True \qquad \alpha \vee \alpha \equiv \alpha$$
$$\alpha \wedge \neg\alpha \equiv False \qquad \alpha \wedge \alpha \equiv \alpha$$

1. $Smoke \Rightarrow Smoke$
2. $(Smoke \Rightarrow Fire) \Rightarrow (\neg Smoke \Rightarrow \neg Fire)$
3. $Smoke \vee Fire \vee \neg Fire$
4. $(Fire \Rightarrow Smoke) \wedge Fire \wedge \neg Smoke$

1. $\equiv \neg Smoke \vee Smoke$
$\equiv True \Rightarrow$ Valid / Satisfiable

2. $\equiv \neg(\neg Smoke \vee Fire) \vee (\neg\neg Smoke \vee \neg Fire)$
$\equiv (\neg\neg Smoke \wedge \neg Fire) \vee (Smoke \vee \neg Fire)$
$\equiv (Smoke \wedge \neg Fire) \vee (Smoke \vee \neg Fire)$
$\equiv (Smoke \vee \neg Fire \vee Smoke) \wedge (Smoke \vee \neg Fire \vee \neg Fire)$
$\equiv (Smoke \vee \neg Fire) \wedge (Smoke \vee \neg Fire)$
$\equiv Smoke \vee \neg Fire$
$\Rightarrow$ not valid / satisfiable

Smoke	Fire	Smoke $\vee \neg$ Fire
T	T	T
T	F	T
F	T	F
F	F	T

3. $\equiv Smoke \vee True$
$\Rightarrow$ valid / satisfiable

4. $\equiv (\neg Fire \vee Smoke) \wedge Fire \wedge \neg Smoke$
$\equiv [(\neg Fire \wedge Fire) \vee (Smoke \wedge Fire)] \wedge \neg Smoke$
$\equiv [(\neg Fire \wedge Fire) \wedge \neg Smoke] \vee [(Smoke \wedge Fire) \wedge \neg Smoke]$
$\equiv [False \wedge \neg Smoke] \vee [False \wedge Fire]$
$\equiv False$
$\Rightarrow$ unsatisfiable

Problem 4.3.1: How can you formalize the sentence “A says that B is a knight”?

Problem 4.3.2: Assume that $Remark$ represents what a person says and that we can represent it using propositional logic[1]. Additionally, assume that P could either be A, B, or C. From the previous problem, can you generalize the method to model the sentence “person P says (or replies) $Remark$”?

Problem 4.3.3: Model the following facts which are taken from the puzzle:

1. B replies, “A said that he is a knave.”
2. C says, “Don’t believe B; he’s lying!”

Problem 4.3.4: By using the following logical equivalences

$$(X \Leftrightarrow \neg X) \equiv \mathit{False}$$
$$(X \Leftrightarrow \mathit{False}) \equiv \neg X$$

and the following deduction (inference) rule

$$\frac{P \Leftrightarrow Q \qquad Q}{P}$$

deduce what B and C are.

Problem 4.4: Superman does not exist

(*Proof by resolution.*) The following text is taken from [3].

> If Superman were able and willing to prevent evil, he would do so. If Superman were unable to prevent evil, he would be impotent; if he were unwilling to prevent evil, he would be malevolent. Superman does not prevent evil. If Superman exists, he is neither impotent nor malevolent. Therefore, Superman does not exist.

Assume that we use the following propositions and their meaning:

A : Superman is able to prevent evil.
W : Superman is willing to prevent evil.
I : Superman is impotent.
M : Superman is malevolent.
P : Superman prevents evil.
E : Superman exists.

Problem 4.4.1: Formalize the facts from the text using the propositions defined above.

Problem 4.4.2: Assume we want to prove that “Superman does not exist” using the resolution approach for propositional logic. Identify which sentences belong to the knowledge base KB, and which sentence we want to deduce. How do we need to process these sentences before applying the resolution principle?

Problem 4.4.3: Prove diagramatically with the resolution approach that “Superman does not exist.”

Problem 4.5: Completeness and soundness

Recall the definition of *completeness* and *soundness*:

[1] This is a mouthful way to say that $Remark$ is a propositional logic sentence.

Problem 4.3: Knights and Knaves

(*Proof by inference rule.*) The following puzzle is taken from [2]. Suppose we are on an island with two types of inhabitants: "knights" who always tell the truth, and "knaves" who always lie.

> According to this problem, three of the inhabitants – A, B and C – were standing together in the garden. A stranger passed by and asked A, "Are you a knight or a knave?". A answered, but rather indistinctly, so the stranger could not make out what he said. The stranger then asked B, "What did A say?". B replied, "A said that he is a knave". At this point the third man, C, said "Don't believe B; he's lying!". The question is, what are B and C?

Model this logic puzzle by introducing three atomic propositions A, B, and C with intended interpretation that A, B, and C are knights.

Problem 4.3.1: How can you formalize the sentence "A says that B is a knight"?

$A \Leftrightarrow B$

Problem 4.3.2: Assume that *Remark* represents what a person says and that we can represent it using propositional logic[1]. Additionally, assume that P could either be A, B, or C. From the previous problem, can you generalize the method to model the sentence "person P says (or replies) *Remark*"?

$P \Leftrightarrow Remark$

Problem 4.3.3: Model the following facts which are taken from the puzzle:

1. B replies, "A said that he is a knave." $\Rightarrow B \Leftrightarrow (A \Leftrightarrow \neg A)$
2. C says, "Don't believe B; he's lying!" $\Rightarrow C \Leftrightarrow \neg B$

Problem 4.3.4: By using the following logical equivalences

$$(X \Leftrightarrow \neg X) \equiv False$$
$$(X \Leftrightarrow False) \equiv \neg X$$

and the following deduction (inference) rule

$$\frac{P \Leftrightarrow Q \qquad Q}{P}$$

deduce what B and C are.

$\equiv B \Leftrightarrow (A \Leftrightarrow \neg A)$

$\equiv B \Leftrightarrow False$

$\equiv \neg B$ knaive

$$\frac{C \Leftrightarrow \neg B \qquad \neg B}{C}$$

$\Rightarrow C$ knight

Problem 4.4: Superman does not exist

(*Proof by resolution.*) The following text is taken from [3].

> If Superman were able and willing to prevent evil, he would do so. If Superman were unable to prevent evil, he would be impotent; if he were unwilling to prevent evil, he would be malevolent. Superman does not prevent evil. If Superman exists, he is neither impotent nor malevolent. Therefore, Superman does not exist.

Assume that we use the following propositions and their meaning:

A : Superman is able to prevent evil.
W : Superman is willing to prevent evil.
I : Superman is impotent.
M : Superman is malevolent.
P : Superman prevents evil.
E : Superman exists.

Problem 4.4.1: Formalize the facts from the text using the propositions defined above.

1. If Superman were able and willing to prevent evil, he would do so.
   $(A \wedge W) \Rightarrow P$
2. If Superman were unable to prevent evil, he would be impotent.
   $\neg A \Rightarrow I$
3. If he were unwilling to prevent evil, he would be malevolent.
   $\neg W \Rightarrow M$
4. Superman does not prevent evil.
   $\neg P$
5. If Superman exists, he is neither impotent nor malevolent.
   $E \Rightarrow (\neg I \wedge \neg M)$

Problem 4.4.2: Assume we want to prove that "Superman does not exist" using the resolution approach for propositional logic. Identify which sentences belong to the knowledge base KB, and which sentence we want to deduce. How do we need to process these sentences before applying the resolution principle?

$\Rightarrow \neg E$

$KB \wedge \neg\alpha \Rightarrow KB \wedge E$

Problem 4.4.3: Prove diagramatically with the resolution approach that "Superman does not exist."

$$(A \wedge W) \Rightarrow P \equiv \neg(A \wedge W) \vee P$$
$$\equiv \neg A \vee \neg W \vee P$$

$$\neg A \Rightarrow I \equiv \neg\neg A \vee I$$
$$\equiv A \vee I$$

$$\neg W \Rightarrow M \equiv \neg\neg W \vee M$$
$$\equiv W \vee M$$

$$\neg P \equiv \neg P$$

$$E \Rightarrow (\neg I \wedge \neg M) \equiv \neg E \vee (\neg I \wedge \neg M)$$
$$\equiv (\neg E \vee \neg I) \wedge (\neg E \vee \neg M)$$

$\neg A \vee \neg W \vee P$ $\quad A \vee I$ $\quad W \vee M$ $\quad \neg P$ $\quad \neg E \vee \neg I$ $\quad \neg E \vee \neg M$ $\quad E$

$\neg A \vee \neg W$

$\neg W \vee I$

$M \vee I$

$\neg E \vee M$

$\neg E \vee \neg E = \neg E$

ϕ

Completeness: An inference algorithm is **complete** if and only if for every entailed sentence $KB \models \alpha$, the inference algorithm will always be able to derive it.

Soundness: An inference algorithm is **sound** if and only if for every sentence it derives, it is guaranteed that the sentence is entailed $KB \models \alpha$.

Problem 4.5.1: Suppose that we have an inference algorithm which will *always* be able to derive a given sentence (regardless whether it is entailed or not). Would this inference algorithm be complete? Sound?

complete but unsound

Problem 4.5.2: Suppose now that we have an inference algorithm which will *never* be able to derive a given sentence (regardless whether it is entailed or not). Would this inference algorithm be complete? Sound?

sound but not complete

References

[1] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2010.

[2] R. Backhouse, *Algorithmic Problem Solving*, 1st. Wiley Publishing, 2011.

[3] D. Gries and F. B. Schneider, *A Logical Approach to Discrete Math*. Springer-Verlag New York, Inc., 1993.