safeAI | checking logical models

ANNELINE DAGGELINCKX, MATTHIJS KEMP, and OTTO MÄTTAS, Utrecht University, The Netherlands

ACM Reference Format:
Anneline Daggelinckx, Matthijs Kemp, and Otto Mättas. 2021. safeAI | checking logical models. 1, 1 (May 2021), 1 page. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 WEEK 3 ASSIGNMENTS

1.1 Describing Kripke models

Below, a description for a distributed system is given in a Kripke model. It is a structure consisting of a certain set of ordinary models for classical logic, ordered by a certain relation, and serving for the interpretation of various non-classical logics (intuitionistic, modal, etc.)

1.1.1 States. First, to define model M_{abc}, all the possible states need to be described. Let the states be S such that $S = \{s0, s1, ..., s5\}$, where

- $s0 = \{a1, b2, c3\}$;
- $s1 = \{a1, b3, c2\}$;
- $s2 = \{a2, b1, c3\}$;
- $s3 = \{a2, b3, c1\}$;
- $s4 = \{a3, b1, c2\}$;
- $s5 = \{a3, b2, c1\}$.

1.1.2 Indistinguishability relations. The states with indistinguishable knowledge for each agent $A = \{a, b, c\}$ have been described as

- $a = \{s0, s1\}, \{s2, s3\}, \{s4, s5\}$;
- $b = \{s2, s4\}, \{s0, s1\}, \{s1, s3\}$;
- $c = \{s3, s5\}, \{s1, s4\}, \{s0, s2\}$.

1.1.3 Valuation. Let the valuations be V such that $V(a_x) = \{S_a\}$, where

- $V(a1) = \{s0, s1\}$;
- $V(a2) = \{s2, s3\}$;
- $V(a3) = \{s4, s5\}$;
- $V(b1) = \{s2, s4\}$;
- $V(b3) = \{s1, s3\}$;
- $V(b2) = \{s0, s5\}$;
- $V(c1) = \{s3, s5\}$;
- $V(c2) = \{s1, s4\}$;
- $V(c3) = \{s0, s2\}$.

Authors' address: Anneline Daggelinckx, a.daggelinckx@students.uu.nl; Matthijs Kemp, m.g.r.kemp@students.uu.nl; Otto Mättas, o.mattas@students.uu.nl, Utrecht University, P.O. Box 80125, Utrecht, Utrecht, The Netherlands, 3508 TC.

XXXX-XXXX/2021/5-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1.2 Proving Kripke models using ELCD

Below, there are three statements to help prove whether the model is properly constructed:

(1) it is distributed knowledge between a, b and c that $a1 \wedge b2 \wedge c3$;
(2) it is common knowledge between a, b and c that $a1 \vee b2 \vee c3$;
(3) it is common knowledge between a, b and c that $a1 \vee a2 \vee a3$.

1.2.1 Translation. First, the statements need to be translated from English to the notation form of epistemic logic with common and distributed knowledge (ELCD):

(1) $D_{a,b,c}(a1 \wedge b2 \wedge c3)$;
(2) $C_{a,b,c}(a1 \vee b2 \vee c3)$;
(3) $C_{a,b,c}(a1 \vee a2 \vee a3)$.

1.2.2 Argumentation. Then, it is possible to determine whether the statement holds (is true) or not:

(1) This statement is true. Each agent knows it's own card. Agent a knows $a1$, agent b knows $b2$ and agent c knows $c3$. Together, the agents would know which agent holds which card.
(2) This statement is false. Agent c knows that b holds card 1 or 2. It can reason that if b has card 2, agent b knows that either situation $a2b1c3$ or $a3b1c2$ holds. In the latter, $(a1 \vee b2 \vee c3)$ would be false. Therefore, agent c doesn't know that agent b knows $(a1 \vee b2 \vee c3)$ and it is not common knowledge.
(3) This statement is true. The agents know that there are only three cards and that every agent has one card. Therefore, they know that agent a has to have a card with a single value $\{1; 2; 3\}$. Since this is defined by the games rules, the agents know that the others also know this. Therefore it is common knowledge.

1.3 Investigating truthfulness in Kripke models

2 WEEK 4 ASSIGNMENTS

2.1 Designing interpreted systems in ISPL

2.2 Designing model checking algorithms

REFERENCES