safeAI | checking logical models

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

1 WEEK 8 ASSIGNMENTS

1.1 Defining Concurrent Epistemic Game Structures

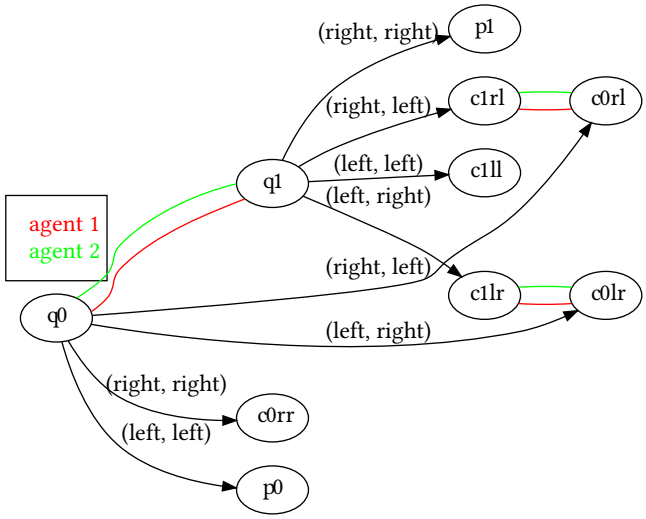
Consider a CEGS $M_{chicken}$, where $(\mathbb{A}gt, St, Act, d, out)$ is a concurrent game structure, and $\sim_a$ are indistinguishability relations over St, one per agent in $\mathbb{A}gt$. We can now define the CEGS as a tuple

$$M_{chicken} = (\mathbb{A}gt, St, \sim_a | a \in \mathbb{A}gt, Act, d, out) \quad (1)$$

, where

- $\mathbb{A}gt = \{a_1, a_2\}$
- $St =$
  - $q0$ =in country where traffic is left
  - $q1$ =in country where traffic is right
  - $q2$ =in country where traffic is left; a1 drives left; a2 drives left
  - $q3$ =in country where traffic is left; a1 drives left; a2 drives right
  - $q4$ =in country where traffic is left; a1 drives right; a2 drives left
  - $q5$ =in country where traffic is left; a1 drives right; a2 drives right
  - $q6$ =in country where traffic is right; a1 drives right; a2 drives right
  - $q7$ =in country where traffic is right; a1 drives right; a2 drives left
  - $q8$ =in country where traffic is right; a1 drives left; a2 drives right
  - $q9$ =in country where traffic is right; a1 drives left; a2 drives left
- $\sim_a = \{\{q_0, q_1\}, \{q_3, q_8\}, \{q_4, q_7\}, \{q_2\}, \{q_5\}, \{q_6\}, \{q_9\}\}$
- $Act = \{drive\_left, drive\_right\}$
- $d(\mathbb{A}gt, q_i) = \{drive\_left, drive\_right\}, \forall i \in \{0, ..., 9\}$
- $out =$
  - $out(q_0, left, left) = q_2$
  - $out(q_0, left, right) = q_3$
  - $out(q_0, right, left) = q_4$
  - $out(q_0, right, right) = q_5$
  - $out(q_1, left, left) = q_9$
  - $out(q_1, left, right) = q_8$
  - $out(q_1, right, left) = q_7$
  - $out(q_1, right, right) = q_6$
- let the evaluations be V, where
  - $V(crash) = \{q_3, q_4, q_5, q_7, q_8, q_9\}$
  - $V(left) = \{q_0, q_2, q_3, q_4, q_5\}$

As there is no proposition for taking action $drive\_right$, we are evaluating q_6 implicitly.



1.2 Validating Concurrent Epistemic Game Structures

For the defined CEGS $M_{chicken}$, it is **untrue** that agent 1 has a memoryless strategy to enforce $\neg crash$ in the next state ($<<1>>X\neg crash$), starting from $q0$ because in q_0, the only way to not crash is if both agents drive left. Agent a_1 itself cannot force this protocol on its own, agent a_2 needs to adhere to it as well.

1.3 3

1.4 4

1.5 5

1.6 6

1.7 7

1.8 8

1.9 9

1.10 10

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.