Logică pentru cunoaștere - material suplimentar

Cursul opțional Elemente de securitate și logică aplicată,

Anul 3, Semestrul 2, 2024

Laurențiu Leuștean

Web page: http://cs.unibuc.ro/~lleustean/

Puzzle-ul cu copiii noroioși (Muddy children puzzle)

- Un grup de n copii intră în casă după ce s-au jucat afară în noroi. Sunt întâmpinați pe hol de tatăl lor, care observă că k dintre copii au noroi pe frunte.
- Tatăl face următorul anunț: Cel puțin unul dintre voi are noroi pe frunte.
- Fiecare copil poate vedea fruntile tuturor celorlalți copii, dar nu și pe a sa.
- Apoi tatăl întreabă: Știe cineva dintre voi că are noroi pe frunte?
- Toți copiii răspund Nu.
- Tatăl repetă întrebarea, din nou toți copiii răspund Nu.
- Tatăl nu cedează și continuă să repete întrebarea.
- După exact k repetiții, toți copiii cu noroi pe frunte răspund Da simultan.

Puzzle-ul cu copiii noroioși

$k = 1$

- Există un singur copil cu noroi pe frunte.
- Acesta știe că ceilalți copii sunt curați.
- Când tatăl zice că cel puțin un copil este noroios, copilul noroios concluzionează că trebuie să fie el.
- Niciunul dintre ceilalți copii nu știe în acest moment dacă este sau nu noroios.
- Copilul noroios răspunde Da la prima întrebare a tatălui.
- După ce acesta răspunde Da, ceilalți copii știu că sunt curați.

Puzzle-ul cu copiii noroioși

$k = 2$

- Există doi copii cu noroi pe frunte.
- Imaginează-ți că tu ești unul din cei doi copii noroioși.
- Vezi că unul din ceilalți copii este noroios.
- După primul anunț al tatălui, nu ai suficiente informații pentru a ști dacă ai noroi pe frunte. S-ar putea să fii noroios, dar se poate și ca celălalt copil să fie singurul cu noroi pe frunte.
- Prin urmare, răspunzi Nu la prima întrebare a tatălui.
- Observi că celălalt copil noroios răspunde Nu.
- Îți dai seama că și tu trebuie să fii noroios, altfel acel copil ar fi răspuns Da.
- Așadar, după a doua întrebare a tatălui, răspunzi Da. Desigur, la fel face și celălalt copil noroios.

Acest argument se poate extinde la $k = 3, 4, \ldots$.

Puzzle-ul cu copiii noroioși

Analizăm puzzle-ul cu copiii noroioși folosind logica epistemică.

Presupunem că următoarele sunt cunoaștere comună:

- tatăl este sincer,
- toți copiii îl aud pe tată,
- fiecare copil vede frunțile celorlalți copii,
- niciun copil nu vede fruntea proprie,
- toți copiii sunt sinceri și extrem de inteligenți.

Puzzle-ul cu copiii noroioși

Presupunem că sunt n copii; îi numerotăm $1, \dots, n$. Așadar, $Ag = \{1, \dots, n\}$.

- Considerăm prima dată situația de dinainte de a vorbi tatăl.
- Câtiva copii sunt noroioși, ceilalți sunt curați.
- Descriem o posibilă situație printr-un n-tuplu de 0 și 1 de forma $(x_1, \dots, x_n)$, unde $x_i = 1$ dacă copilul i este noroios și $x_i = 0$ altfel.
- Există 2^n situații posibile.

Puzzle-ul cu copiii noroioși

$n = 3$

- Să presupunem că situația actuală este descrisă de tuplul $(1, 0, 1)$, care spune că copilul 1 și copilul 3 sunt noroioși, în timp ce copilul 2 este curat.
- Ce situații consideră posibile copilul 1 înainte ca tatăl să vorbească?
- Deoarece copilul 1 poate vedea frunțile celorlalți copii dar nu vede propria frunte, singura lui îndoială este dacă el este noroios sau curat. Prin urmare, copilul 1 consideră două situații posibile: $(1, 0, 1)$ (situația actuală) și $(0, 0, 1)$. Similar, copilul 2 consideră două situații posibile: $(1, 0, 1)$ și $(1, 1, 1)$.

În general, copilul i are aceeași informație în două situații posibile (x_1, x_2, x_3) și (y_1, y_2, y_3) dacă și numai dacă $x_j = y_j$ pentru orice $j \neq i$.

Puzzle-ul cu copiii noroioși

Putem descrie situația generală cu ajutorul cadrului

$$\mathcal{F} = (W, \mathcal{K}_1, \dots, \mathcal{K}_n),$$

unde

- $W = \{(x_1, \dots, x_n) \mid x_i \in \{0, 1\} \text{ pentru orice } i = 1, \dots, n\}$. Deci W are 2^n stări.
- Pentru orice $i = 1, \dots, n$,

  $\mathcal{K}_i wv$ ddacă w și v coincid pe toate componentele cu posibila excepție a componentei i.
- Se poate verifica ușor ca relațiile $\mathcal{K}_i$ sunt de echivalență.

Prin urmare, $\mathcal{F}$ este un cadru pentru logica epistemică ***S*5**.

Mai trebuie să definim $PROP$ și evaluarea $V : PROP \to 2^W$.

- Deoarece vrem să raționăm despre faptul că un anumit copil este noroios sau nu, definim $PROP = \{p_1, \ldots, p_n, p\}$, unde, intuitiv, p_i înseamnă copilul i este noroios, în timp ce p înseamnă cel puțin un copil este noroios.
- Definim V astfel:

  $V(p_i) = \{(x_1, \ldots, x_n) \in W \mid x_i = 1\}$,

  $V(p) = \{(x_1, \ldots, x_n) \in W \mid \text{ există } j = 1, \ldots, n \text{ a.î. } x_j = 1\}$.
- Rezultă că

  $\mathcal{M}, (x_1, \ldots, x_n) \Vdash p_i$ ddacă $x_i = 1$,

  $\mathcal{M}, (x_1, \ldots, x_n) \Vdash p$ ddacă există $j = 1, \ldots, n$ a.î. $x_j = 1$.

Avem un model cu 2^n noduri, fiecare descris de un n-tuplu de 0 și 1, astfel încât două noduri sunt unite printr-o muchie ddacă sunt diferite în cel mult o componentă.

Reamintim că omitem buclele și săgețile arcelor.

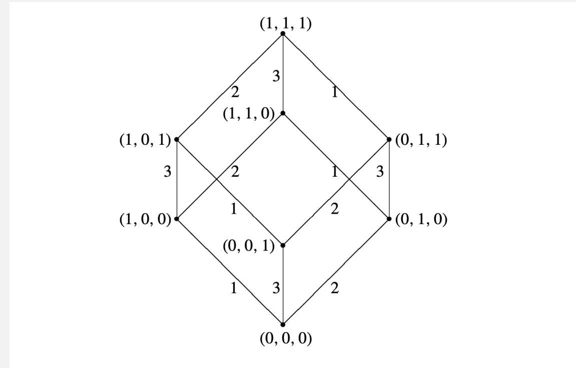
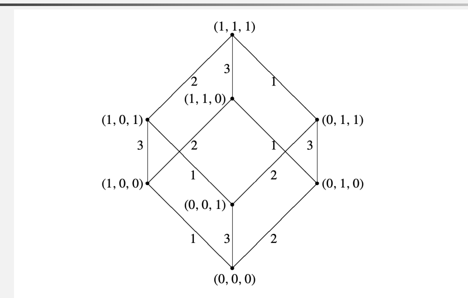


Figura 1: Cadru pentru puzzle-ul cu $n = 3$

- $\mathcal{M}, (1, 0, 1) \Vdash K_1 \neg p_2$;
- $\mathcal{M}, (1, 0, 1) \Vdash K_1 p_3$;
- $\mathcal{M}, (1, 0, 1) \Vdash \neg K_1 p_1$.

Studiem în continuare ce se întâmplă după ce vorbește tatăl.

- Tatăl spune p, care este deja cunoscută de toți copiii dacă avem cel puțin doi copii noroioși.
- Totuși, starea cunoașterii se schimbă.



- În $(1, 0, 1)$, copilul 1 consideră situația $(0, 0, 1)$ posibilă și în $(0, 0, 1)$ copilul 3 consideră $(0, 0, 0)$ posibilă.
- În $(1, 0, 1)$, înainte ca tatăl să vorbească, chiar dacă toți știu că cel puțin un copil este noroios, copilul 1 crede că este posibil ca copilul 3 să creadă că e posibil ca niciun copil să fie noroios.
- După ce vorbește tatăl, devine cunoaștere comună că cel puțin un copil este noroios.

Puzzle-ul cu copiii noroioși

- În cazul general, reprezentăm grafic schimbarea în starea de cunoaștere a grupului prin eliminarea punctului $(0, 0, \ldots, 0)$ din cub.
- Mai precis, ce se întâmplă este că nodul $(0, 0, \ldots, 0)$ rămâne, dar toate muchiile între $(0, 0, \ldots, 0)$ și nodurile cu exact un 1 dispar, deoarece este cunoaștere comună că, chiar dacă un singur copil e noroios, după ce vorbește tatăl, acel copil nu va considera posibilă situația în care niciun copil nu este noroios.

Puzzle-ul cu copiii noroioși

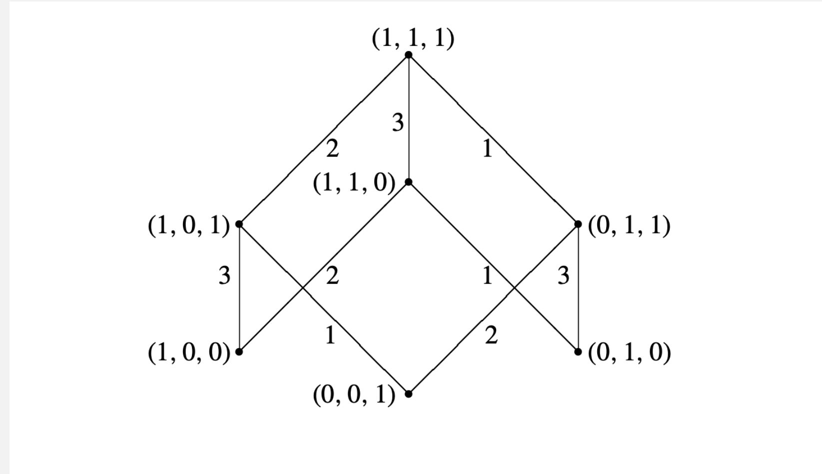


Figura 2: Cadru pentru $n = 3$ după ce vorbește tatăl

Puzzle-ul cu copiii noroioși

De fiecare dată când copiii răspund Nu la întrebarea tatălui, starea de cunoaștere a grupului se modifică și cubul este trunchiat în continuare.

- Să considerăm ce se întâmplă după ce copiii răspund Nu la prima întrebare a tatălui.
- Acum toate nodurile cu exact un 1 pot fi eliminate. Mai precis, muchiile între aceste noduri și nodurile cu doi 1 dispar.
- Nodurile cu cel mult un 1 nu mai sunt accesibile din noduri cu doi sau mai mulți 1.

Puzzle-ul cu copiii noroioși

- Dacă situația actuală este descrisă, să zicem, de tuplul $(1, 0, \ldots, 0)$, atunci copilul 1 consideră inițial două situații posibile: $(1, 0, \ldots, 0)$, și $(0, 0, \ldots, 0)$.
- Deoarece după ce vorbește tatăl devine cunoaștere comună că $(0, 0, \ldots, 0)$ nu este posibilă, el știe atunci că situația este descrisă de $(1, 0, \ldots, 0)$. Prin urmare, știe că este noroios.
- După ce toți copiii răspund Nu la prima întrebare a tatălui, este cunoaștere comună că situația nu poate fi $(1, 0, \ldots, 0)$.
- Raționamente similare ne permit să eliminăm orice situație cu exact un 1.
- Deci, după ce toți copiii răspund Nu la prima întrebare a tatălui, este cunoaștere comună că cel puțin doi copii sunt noroioși.

Puzzle-ul cu copiii noroioși

- Argumente de acest fel pot fi folosite pentru a arăta că după ce copiii răspund Nu de m ori, putem elimina toate nodurile cu cel mult m de 1 (sau, mai precis, deconectăm aceste noduri de restul grafului).
- Avem astfel un șir de cadre care descriu cunoașterea copiilor la fiecare pas din proces.
- În esență, ceea ce se întâmplă este că, dacă, într-un nod w, devine cunoscut că un nod v este imposibil, atunci pentru fiecare nod u accesibil din w, muchia dintre u și v (dacă există una) este eliminată.

Puzzle-ul cu copiii noroioși

- După m runde, este cunoaștere comună că cel puțin $m+1$ copii sunt noroioși.
- Dacă situația reală este descrisă de un tuplu cu exact $m+1$ de 1, atunci înainte ca tatăl să întrebe pentru a $(m+1)$-a dată, copiii noroioși vor ști situația exactă și, în particular, vor ști că sunt noroioși și, ca urmare, vor răspunde Da.
- Ei nu puteau să răspundă Da mai devreme, deoarece până la acest moment, fiecare copil murdar considera posibil cazul în care e curat.

Puzzle-ul cu copiii noroioși

- Conform modului în care modelăm cunoașterea în acest context, un copil știe un fapt dacă acel fapt rezultă din informația sa curentă.
- Dacă unul dintre copii nu ar fi prea inteligent, atunci nu ar putea să realizeze că știe că este noroios, chiar dacă ar avea în principiu destulă informație pentru a ști acest lucru.
- Pentru a răspunde Da la întrebarea tatălui, copilul trebuie să cunoască ce se poate deduce din informația pe care o are.
- Presupunem implicit că (este cunoaștere comună că) toți copiii sunt logic omniscienți, adică ei sunt destul de inteligenți pentru a calcula toate consecințele informației pe care o au.
- Această omniștiință logică este cunoaștere comună.

Puzzle-ul cu copiii noroioși

- Să considerăm acum situația în care tatăl nu spune inițial p.
- Înainte ca tatăl să vorbească, situația este descrisă de cubul n-dimensional.
- Când tatăl pune prima dată întrebarea Știe cineva dintre voi că are noroi pe frunte?, copiii răspund Nu, deoarece în orice situație, fiecare copil consideră posibilă situația în care este curat.
- Nu se câștigă nicio informație din acest răspuns, așa că situația este reprezentată de cubul n-dimensional.
- Se poate arăta prin inducție după m că este cunoaștere comună că la a m-a întrebare a tatălui toți copiii răspund Nu și starea de cunoaștere după a m-a întrebare a tatălui este descrisă tot de cub.
- Prin urmare, starea de cunoaștere a copiilor nu se schimbă, indiferent de numărul de întrebări ale tatălui.

Modelele cu partiții pentru cunoaștere (*partition models of knowledge*) sunt definite în

Yoav Shoham, Kevin Leyton-Brown, Multiagents Systems, Cambridge University Press, 2009.

Reamintim că $n \geq 1$ și $AG = \{1, \ldots, n\}$ este mulțimea agenților.

Definiția 0.1

Un cadru cu partiții (partition frame) este un tuplu $\mathcal{P}_F = (W, I_1, \ldots, I_n)$, *unde*

- W *este o mulțime nevidă de lumi posibile.*
- *Pentru orice* $i = 1, \ldots, n$, I_i *este o partiție a lui* W.

Ideea este că I_i partiționează W în mulțimi de lumi posibile care sunt imposibil de distins (*indistinguishable*) din punctul de vedere al agentului i.

Fie A o mulțime nevidă.

Recapitulare: O partiție a lui A este o familie $(A_j)_{j \in J}$ de submulțimi nevide ale lui A care verifică proprietățile:

$$A = \bigcup_{j \in J} A_j \text{ și } A_j \cap A_k = \emptyset \text{ pentru orice } j \neq k.$$

Recapitulare: Există o bijecție între mulțimea relațiilor de echivalență pe A si mulțimea partițiilor lui A:

- $(A_j)_{j \in J}$ partiție a lui $A \mapsto$ relația de echivalență pe A definită prin: $x \sim y \Leftrightarrow$ există $j \in J$ astfel încât $x, y \in A_j$.
- $\sim$ relație de echivalență pe $A \mapsto$ partiția care constă în toate clasele de echivalență diferite ale lui $\sim$.

Fie $i = 1, \ldots, n$,

- Notăm $\mathcal{K}_i$ relația de echivalență corespunzătoare partiției I_i.
- Pentru orice lume $w \in W$, fie $I_i(w)$ clasa de echivalență a lui w în relația $\mathcal{K}_i$. $I_i(w)$ este mulțimea lumilor posibile pe care agentul i nu le poate distinge de w.
- Pentru orice lumi $v, w \in W$, $v \in I_i(w)$ ddacă $\mathcal{K}_i wv$ ddacă agentul i nu poate distinge v de w.

$\mathcal{F} = (W, \mathcal{K}_1, \ldots, \mathcal{K}_n)$ este un cadru Kripke pentru logica epistemică ***S*5**.

Cadru cu partiții = cadru Kripke pentru logica epistemică ***S*5**

Definiția 0.2

Un model cu partiții (partition model) peste un limbaj Σ *este un tuplu* $\mathcal{P}_M = (\mathcal{P}_F, \pi)$, *unde*

- $\mathcal{P}_F = (W, I_1, \ldots, I_n)$ *este un cadru cu partiții.*
- $\pi : \Sigma \to 2^W$ *este o funcție de interpretare.*

Pentru orice afirmație $\varphi \in \Sigma$, gândim $\pi(\varphi)$ ca fiind mulțimea lumilor posibile din modelul cu partiții $\mathcal{P}_M$ unde φ este satisfăcută.

Putem lua, de exemplu, Σ ca fiind o mulțime de propoziții atomice sau o mulțime de formule din logica propozițională.

Folosim notația $K_i\varphi$ pentru "agentul i știe că φ".

Definim în continuare ce înseamnă că o afirmație este adevărată într-un model cu partiții.

Definiția 0.3

Fie $\mathcal{P}_M = (W, I_1, \ldots, I_n, \pi)$ un model cu partiții peste Σ și $w \in W$. Definim relația $\models$ (logical entailment) astfel:

- *Pentru orice $\varphi \in \Sigma$, spunem că $\mathcal{P}_M, w \models \varphi$ ddacă $w \in \pi(\varphi)$.*
- *$\mathcal{P}_M, w \models K_i\varphi$ ddacă pentru orice lume $v \in W$, $v \in I_i(w)$ implică $\mathcal{P}_M, v \models \varphi$.*

Model cu partiții = model Kripke pentru logica epistemică ***S5***.

Putem raționa riguros despre cunoaștere în termeni de modele cu partiții, deci folosind logica epistemică.

Aplicăm modelul cu partiții pentru cunoaștere la puzzle-ul cu copii noroioși (*Muddy children puzzle*).

- Considerăm cazul $n = k = 2$ (doi copii, amândoi noroioși).
- Avem două propoziții atomice: muddy1 și muddy2.
- Există patru lumi posibile:

  w_1 : $muddy1 \wedge muddy2$ (lumea reală)
  w_2 : $muddy1 \wedge \neg muddy2$
  w_3 : $\neg muddy1 \wedge muddy2$
  w_4 : $\neg muddy1 \wedge \neg muddy2$.
- Prin urmare, $\pi(muddy1) = \{w_1, w_2\}$ și $\pi(muddy2) = \{w_1, w_3\}$.
- Sunt două partiții I_1 și I_2.

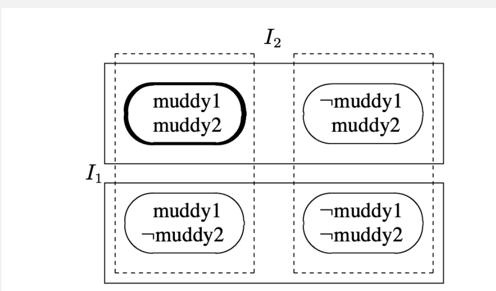
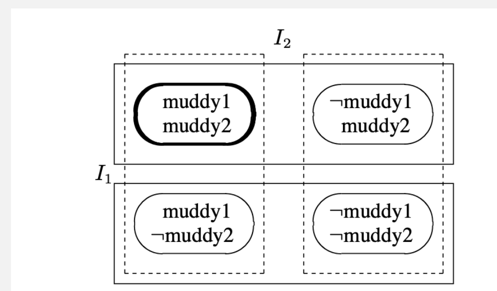


Figura 3: Modelul cu partiții după ce copiii se văd

- Ovalele ilustrează cele patru lumi posibile. Ovalul îngroșat indică adevărata stare a lumii.
- Dreptunghiurile solide reprezintă clasele de echivalență din I_1:
  $I_1(w_1) = I_1(w_3) = \{w_1, w_3\}, \quad I_1(w_2) = I_1(w_4) = \{w_2, w_4\}$
- Dreptunghiurile punctate reprezintă clasele de echivalență din I_2:
  $I_2(w_1) = I_2(w_2) = \{w_1, w_2\}, \quad I_2(w_3) = I_2(w_4) = \{w_3, w_4\}$.



În lumea reală w_1,

- $K_1 muddy2$ și $K_2 muddy1$ sunt adevărate.
- $K_1 muddy1$ nu este adevărată.
- $K_2 muddy2$ nu este adevărată.

Modelul cu partiții pentru cunoaștere

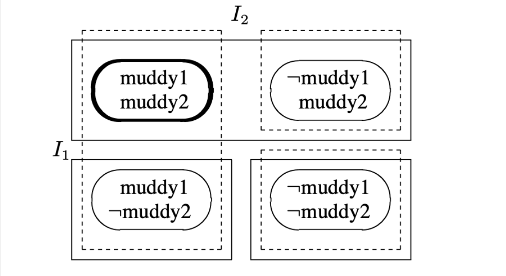


Figura 4: Modelul cu partiții după anunțul tatălui

- Lumea în care niciun copil nu este murdar este eliminată. Starea cunoașterii este prezentată în Figura 4:
  $I_1(w_1) = I_1(w_3) = \{w_1, w_3\}, \quad I_1(w_2) = \{w_2\}, \quad I_1(w_4) = \{w_4\}$
  $I_2(w_1) = I_2(w_2) = \{w_1, w_2\}, \quad I_2(w_3) = \{w_3\}, \quad I_2(w_4) = \{w_4\}.$
- În lumea reală w_1,
  - $K_1 muddy1$ nu este adevărată.
  - $K_2 muddy2$ nu este adevărată.

Modelul cu partiții pentru cunoaștere

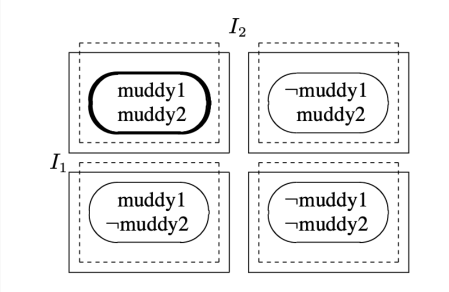


Figura 5: Modelul cu partiții final

- Fiecare copil observă că celălalt copil nu ridică mâna după anunțul tatălui. Prin urmare, fiecare își dă seama că trebuie să fie noroios.
- Starea cunoașterii este prezentată în Figura 5:
  $I_k(w_i) = \{w_i\}$ for all $k = 1, 2,\ i = 1, \ldots, 4.$
- Atât $K_1 muddy1$ cât și $K_2 muddy2$ sunt satisfăcute în lumea reală w_1.