PA054: Formální modely v systémové biologii

David Šafránek

22.3.2013

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

EVROPSKÁ UNIE

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Obsah

Analýza Petriho sítí

Dynamická analýza

Graf dosažitelnosti

Nechť $\mathcal{N} = \langle P, N, f, m_0 \rangle$ je Petriho síť. Graf dosažitelnosti *reach*$(\mathcal{N})$ je difinován jako graf *reach*$(\mathcal{N}) = (V, E)$, kde:

- $V = [m_0\rangle$,
- $E = \{(m, t, m') | m, m' \in [m_0\rangle,\ t \in T.\ m[t\rangle m'\}$.

- k rozhodnutí behaviorálních vlastností, které nelze rozhodnout staticky, je nutné zkonstruovat graf dosažitelnosti
  - pro neohraničnou síť může být nekonečný
  - ohraničenost lze vždy rozhodnout konstrukcí stromu pokrytelnosti (viz IA023)

Analýza Petriho sítí

Dynamická analýza

Nechť $\mathcal{N} = \langle P, N, f, m_0 \rangle$ je Petriho síť t.ž. $reach(\mathcal{N})$ je konečný. Pak platí:

- $\mathcal{N}$ je k-ohraničená, pokud $\forall m \in V_{\mathcal{N}}, p \in P.\, m(p) \leq k$
- $\mathcal{N}$ je reversibilní, pokud $reach(\mathcal{N})$ je silně souvislý
- $\mathcal{N}$ je slabě živá, pokud $\forall m \in V_{\mathcal{N}}, \exists m' \in V_{\mathcal{N}}.\, \langle m, m' \rangle \in E_{\mathcal{N}}$

Analýza Petriho sítí

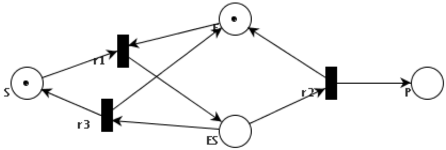
Dynamická analýza – živost sítě

Nechť $\mathcal{N} = \langle P, N, f, m_0 \rangle$ je Petriho síť t.ž. $reach(\mathcal{N})$ je konečný.

1. najdi všechny silně souvislé komponenty (SCC) grafu $reach(\mathcal{N})$ (maximální silně souvislé podgrafy)
2. označ všechny terminální SCC (t.ž. nelze dosáhnout další SCC)
3. `if` $\exists t \in T$ t.ž. není zahrnut v nějaké terminální SCC
   `then` $\mathcal{N}$ není živá
   `else` $\mathcal{N}$ je živá

Analýza Petriho sítí

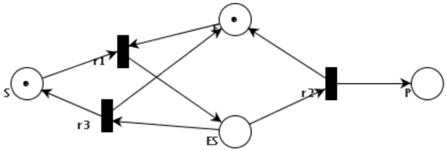
Dynamická analýza – specifické vlastnosti dynamiky



- místo E nikdy nezůstane trvale vyprázdněno
  enzym E je nevyčerpatelný
- iniciální obsah místa S je trvale přesunut do P
  všechny molekuly substrátu S jsou přeměněny v produkt P

Analýza Petriho sítí

Dynamická analýza – specifické vlastnosti dynamiky



- místo E nikdy nezůstane trvale vyprázdněno
  enzym E je nevyčerpatelný
  $\mathbf{GF}(E > 0)$
- iniciální obsah místa S je trvale přesunut do P
  všechny molekuly substrátu S jsou přeměněny v produkt P
  $S == 5 \Rightarrow \mathbf{FG}(P == 5 \wedge S == 0)$

Kripkeho struktura

Nechť AP je množina atomických propozic (obecně boolovské výrazy nad proměnnými, konstantami a predikátovými symboly). *Kripkeho strukturou* nazýváme čtveřici $K = \langle S, S_0, T, L \rangle$ kde:

- S je konečná množina stavů
- $S_0 \subseteq S$ je množina počátečních stavů
- $T \subseteq S \times S$ t.ž. $\forall s \in S, \exists s' \in S : \langle s, s' \rangle \in T$
- L je přiřazení propozic (tzv. interpretační funkce) $L : S \to 2^{AP}$

Kripkeho struktura – vlastnosti

- stav s je *d*eadlockovaný pokud z něj existuje pouze přechod $s \to s$
- pro daný stav $s \in S$ je $L(s)$ množina všech atomických propozic platných v s
- rozbalením Kripkeho struktury z množiny iniciálních stavů je vždy nekonečný strom
    - cesty v tomto stromu představují individuální simulace (běhy) modelovaného systému

Lineární temporální logika – syntax

Nechť AP je množina atomických propozic. Formule φ je formulí *lineární temporální logiky (LTL)* pokud splňuje následující:

- $\varphi = p$ pro libovolné $p \in AP$
- Jsou-li φ_1 a φ_2 formule LTL, pak:
    - $\neg\varphi_1$, $\varphi_1 \wedge \varphi_2$ a $\varphi_1 \vee \varphi_2$ jsou formule LTL
    - $\mathbf{X}\varphi_1$, $\mathbf{F}\varphi_1$ a $\mathbf{G}\varphi_1$ jsou formule LTL
    - $\varphi_1 \mathbf{U} \varphi_2$ a $\varphi_1 \mathbf{R} \varphi_2$ jsou formule LTL

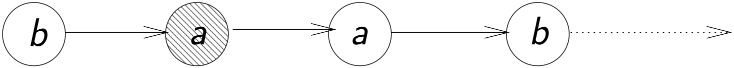
Lineární temporální logika – sémantika

Nechť $\pi = s_0, s_1, \ldots, s_i, \ldots$ je nekonečná posloupnost stavů (běh) v Kripkeho struktuře K. Pro $j > 0$ označme π^j sufix $s_j, s_{j+1}, \ldots, s_i, \ldots$. Definujeme induktivně relaci splnitelnosti $\models$:

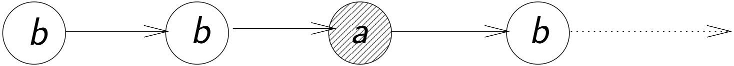
- $\pi \models p$, pokud $p \in L(s_0)$
- $\pi \models \neg\varphi$, pokud $\pi \not\models \varphi$
- $\pi \models \varphi_1 \wedge \varphi_2$, pokud $\pi \models \varphi_1$ a $\pi \models \varphi_2$
- $\pi \models \varphi_1 \vee \varphi_2$, pokud $\pi \models \varphi_1$ nebo $\pi \models \varphi_2$
- $\pi \models \mathbf{X}\varphi$, pokud $\pi^1 \models \varphi$
- $\pi \models \mathbf{F}\varphi$, pokud $\exists i \geq 0.\, \pi^i \models \varphi$
- $\pi \models \mathbf{G}\varphi$, pokud $\forall i \geq 0.\, \pi^i \models \varphi$
- $\pi \models \varphi_1 \mathbf{U} \varphi_2$, pokud $\exists j \geq 0.\, \pi^j \models \varphi_2$ a $\forall i < j.\, \pi^i \models \varphi_1$
- $\pi \models \varphi_1 \mathbf{R} \varphi_2$, pokud $\forall j \geq 0, \forall 0 \leq i < j.\, \pi^i \not\models \varphi_1 \Rightarrow \pi^j \models \varphi_2$.

Lineární temporální logika – sémantika

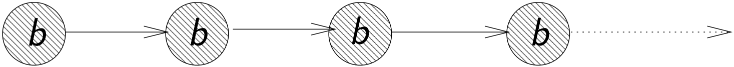
$\mathbf{X}a$



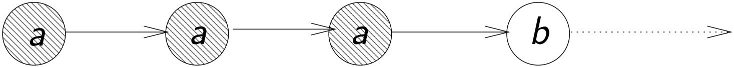
$\mathbf{F}a$



$\mathbf{G}b$



$a\mathbf{U}b$



Lineární temporální logika – sémantika

Pro libovolné formule φ_1, φ_2 platí:
$\neg\mathbf{F}\varphi \equiv \mathbf{G}\neg\varphi$
$\neg(\varphi_1\mathbf{U}\varphi_2) \equiv \neg\varphi_1\mathbf{R}\neg\varphi_2$

K plné expresivitě LTL stačí operátory $\neg, \wedge, \mathbf{X}, \mathbf{U}$.

Lineární temporální logika – sémantika

Pro libovolné formule φ_1, φ_2 platí:
$\neg \mathbf{F}\varphi \equiv \mathbf{G}\neg\varphi$
$\neg(\varphi_1 \mathbf{U} \varphi_2) \equiv \neg\varphi_1 \mathbf{R} \neg\varphi_2$

K plné expresivitě LTL stačí operátory $\neg, \wedge, \mathbf{X}, \mathbf{U}$.

Formule LTL logiky jsou typicky univerzálně interpretovány na Kripkeho struktuře:

Lineární temporální logika – sémantika

Pro libovolné formule φ_1, φ_2 platí:
$\neg\mathbf{F}\varphi \equiv \mathbf{G}\neg\varphi$
$\neg(\varphi_1\mathbf{U}\varphi_2) \equiv \neg\varphi_1\mathbf{R}\neg\varphi_2$

K plné expresivitě LTL stačí operátory $\neg, \wedge, \mathbf{X}, \mathbf{U}$.

Formule LTL logiky jsou typicky univerzálně interpretovány na Kripkeho struktuře:

Nechť K Kripkeho struktura. Řekneme, že formule φ je splněna v K, $K \models \varphi$, pokud pro každý běh $\pi = s_0, \ldots$ t.ž. $s_0 \in S_0$ platí $\pi \models \varphi$.

Model checking

Algoritmus, který pro danou Kripkeho strukturu K a temporální vlastnost Φ rozhodne zda-li $K \models \Phi$. V negativním případě vrátí příklad běhu π t.ž. $\pi \not\models \Phi$.

Logika větvícího se času – CTL

- temporální operátory nahrazeny kvantifikacemi přes nekonečné podstromy
- z hlediska expresivity nesrovnatelná s LTL
- umožňuje zachytit vlastnosti nedeterminismu
- neumožňuje vyjádřit některé temporální vlastnosti

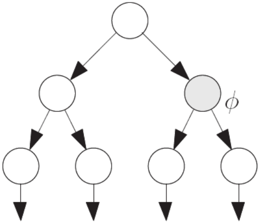
Logika větvícího se času – syntax

Nechť AP je množina atomických propozic. Formule φ je formulí *logiky větvícího se času (CTL)* pokud splňuje následující:

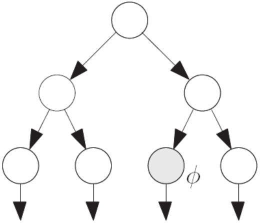
- $\varphi = p$ pro libovolné $p \in AP$
- Jsou-li φ_1 a φ_2 formule CTL, pak:
  - $\neg\varphi_1$, $\varphi_1 \wedge \varphi_2$ a $\varphi_1 \vee \varphi_2$ jsou formule CTL
  - $\mathbf{AX}\varphi_1$, $\mathbf{AF}\varphi_1$ a $\mathbf{AG}\varphi_1$ jsou formule CTL
  - $\mathbf{EX}\varphi_1$, $\mathbf{EF}\varphi_1$ a $\mathbf{EG}\varphi_1$ jsou formule CTL
  - $\mathbf{A}(\varphi_1\mathbf{U}\varphi_2)$ a $\mathbf{E}(\varphi_1\mathbf{U}\varphi_2)$ jsou formule CTL

Pozn.: K plné expresivitě CTL stačí operátory $\neg, \wedge, \mathbf{AX}, \mathbf{AU}, \mathbf{EU}$.

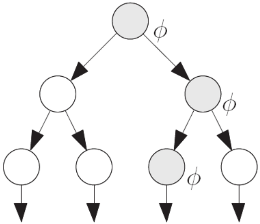
Logika větvícího se času – sémantika



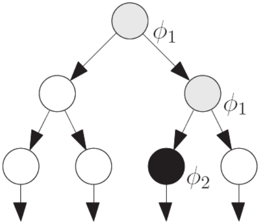
EX ϕ



EF ϕ

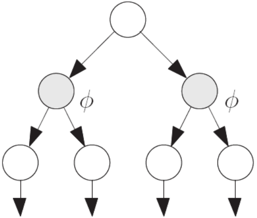


EG ϕ

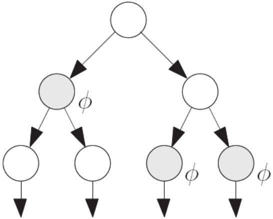


E ϕ_1 **U** ϕ_2

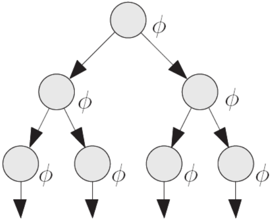
Logika větvícího se času – sémantika



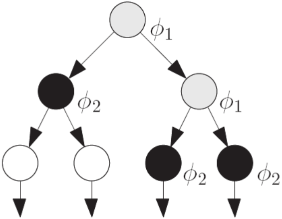
AX ϕ



AF ϕ



AG ϕ



A ϕ_1 **U** ϕ_2

LTL vs. CTL

- oscilace energie mezi volnou a vázanou formou enzymu

$$LTL : \mathbf{G}(((E \wedge \neg ES) \Rightarrow \mathbf{F}(\neg E \wedge ES)) \wedge ((\neg E \wedge ES) \Rightarrow \mathbf{F}(E \wedge \neg ES)))$$

- existence dvou různých stabilních stavů proměnné X a jejich dosažení z iniciálního stavu

$$CTL : \ \mathbf{EFAG}(X > 2) \wedge \mathbf{EFAG}(X \leq 2)$$

Clarke,E.M. and Draghicescu,I.A. (1988) Expressibility results for linear-time and branching-time logics. In Proceedings of REX Workshop, Lecture Notes in Computer Science Vol. 354, Springer, pp. 428–437.

Obsah

Stochastická sémantika modelu

Uvažujme model $\mathcal{M} = \langle S, R, \mathrm{reanet}, \emptyset, \mathrm{map} \rangle$. Označme $n = |S|$.

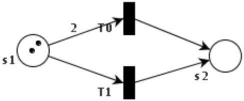
- $\mathrm{SVal} = \mathbb{N}$ ... množství (počet molekul)

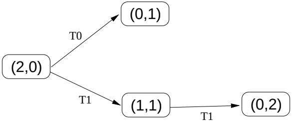
Vektor $[\vec{s}]_{\mathcal{M}} \in \mathrm{SVal}^n$ reprezentuje stav modelu (ohodnocení proměnných) v daném okamžiku.

Ohodnocení vektoru proměnných $\vec{s}$ lze samplovat prostřednictvím vektorové náhodné proměnné $X_{\mathcal{M}}$, $X_{\mathcal{M}} = \langle X_1, ..., X_n \rangle$, kde $X_i : \mathrm{SVal}^n \to \mathbb{N}$ charakterizuje počet molekul substance $s_i \in S$ v daném okamžiku.

Stochastickou (pravděpodobnostní) sémantiku modelu $\mathcal{M}$ definujeme jako stochastický proces $[\mathcal{M}] = \{X_{\mathcal{M}}(t) | t \in \mathbb{R}_0^+\}$.

Pravděpodobnostní funkce označkování



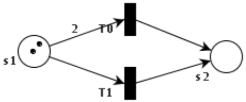


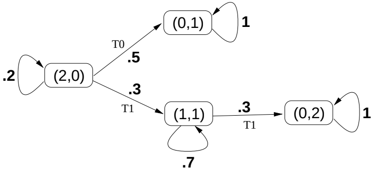
$[\vec{s}]_{\mathcal{M}}$		$X_{\mathcal{M}}$	
s_1	s_2	X_1	X_2
2	0	2	0
1	1	1	1
0	2	0	2
0	1	0	1

- $\Pr\{X_{\mathcal{M}} = (2,0)\} = \Pr\{X_{\mathcal{M}} = (1,1)\} = \Pr\{X_{\mathcal{M}} = (0,2)\} = \Pr\{X_{\mathcal{M}} = (0,1)\} = \frac{1}{4}$
- $\Pr\{X_1 = 1\} = \Pr\{X_1 = 2\} = \Pr\{X_2 = 0\} = \Pr\{X_2 = 2\} = \frac{1}{4}$
- $\Pr\{X_1 = 0\} = \Pr\{X_2 = 1\} = \frac{2}{4} = \frac{1}{2}$
- pravděpodobnostní funkce:

$$p((x_1, x_2)) = \Pr\{X_{\mathcal{M}} = (x_1, x_2)\} = \Pr\{X_1 = x_1 \cap X_2 = x_2\} =$$
$$= \Pr\{X_1 = x_1\} \cdot \Pr\{X_2 = x_2 | X_1 = x_1\}$$

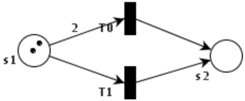
Diskrétní Markovův řetězec a Markovův proces

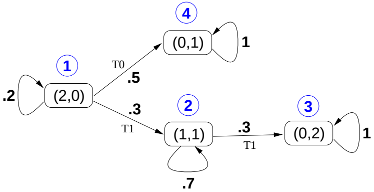




- sledujeme časovou progresi náhodné proměnné $X_{\mathcal{M}}(t)$
- čas uvažujeme jako spočetnou veličinu
- vývoj $X_{\mathcal{M}}$ v čase lze popsat grafem $MC = \langle V, E, p \rangle$:
  - stavy V reprezentují prvky jevového pole
  - přechody E jsou ohodnoceny pravděpodobností $p : E \rightarrow (0, 1\rangle$ t.ž. $\forall v \in V. \sum_{v' \in V} p(\langle v, v' \rangle) = 1$

Diskrétní Markovův řetězec a Markovův proces

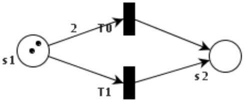


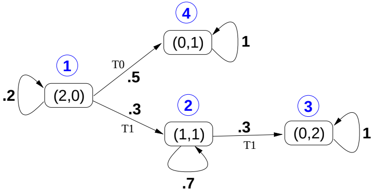


- sledujeme časovou progresi náhodné proměnné $X_{\mathcal{M}}(t)$
- čas uvažujeme jako spočetnou veličinu
- vývoj $X_{\mathcal{M}}$ v čase lze popsat grafem $MC = \langle V, E, p\rangle$:
  - stavy V reprezentují prvky jevového pole
  - přechody E jsou ohodnoceny pravděpodobností $p : E \to (0,1\rangle$ t.ž. $\forall v \in V. \sum_{v' \in V} p(\langle v, v'\rangle) = 1$
  - ekvivalentně lze reprezentovat přechodovou maticí:

$$P = \begin{pmatrix} .2 & .3 & 0 & .5 \\ 0 & .7 & .3 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad P_{ij} = \Pr\{X_{\mathcal{M}}(t+1) = j | X_{\mathcal{M}}(t) = i\}$$

Diskrétní Markovův řetězec a Markovův proces





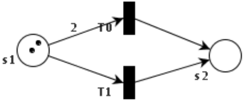
Pravděpodobnost stavu i v čase t budeme značit $p_i(t) = \Pr\{X_{\mathcal{M}}(t) = i\}$. Pro okamžik t máme vektorové rozložení $p(t) = \langle p_1(t), p_2(t), p_3(t), p_4(t)\rangle$.

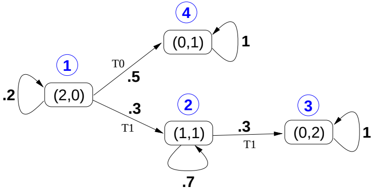
Iniciálně předpokládejme $p_1(0) = p_2(0) = p_3(0) = p_4(0) = \frac{1}{4}$.

Vývoj pravděpodobnostního rozložení proměnné $X_{\mathcal{M}}$ v čase:

$$\begin{aligned} p(1) &= p(0)P \\ p(2) &= p(1)P \\ &\vdots \\ p(k) &= p(0)P^k \end{aligned}$$

Diskrétní Markovův řetězec a Markovův proces





Stochastický proces $\{X_{\mathcal{M}}(t) | t \in \mathbb{N}\}$ rozvíjený dle Markovova řetězce z předchozího slidu se nazývá *Markovův proces*.

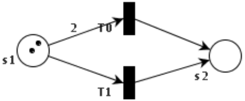
Klíčovou vlastností Markovova procesu je nezávislost na historii (tzv. "memoryless"):

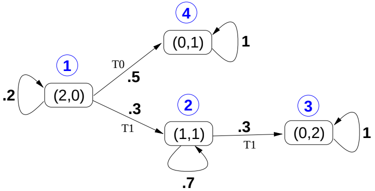
Pravděpodobnostní rozložení proměnné $X_{\mathcal{M}}(t+1)$ *závisí pouze na bezprostředně (v čase) předchozím rozložení* $X_{\mathcal{M}}(t)$.

$$p_j(t+1) = \sum_{i=1}^{k} P_{ij} p_i(t)$$

Diskrétní Markovův řetězec a Markovův proces

Příklad



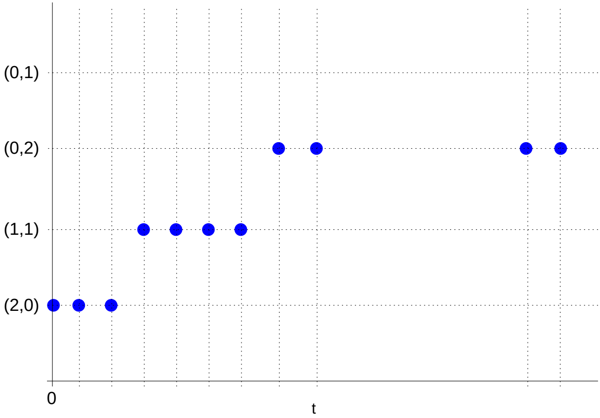


	1	2	3	4
$p(0)$:	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
$p(1)$:	0.0500	0.2500	0.3250	0.3750
$p(2)$:	0.0100	0.1900	0.4000	0.4000
$p(3)$:	0.0020	0.1360	0.4570	0.4050
$p(4)$:	0.0004	0.0958	0.4978	0.4060
$p(5)$:	0.0001	0.0672	0.5265	0.4062
$p(6)$:	0.0000	0.0470	0.5467	0.4062
$p(7)$:	0.0000	0.0329	0.5608	0.4062
$p(8)$:	0.0000	0.0231	0.5707	0.4062
		$\vdots$		
$p(26)$:	0.0000	0.0000	0.5937	0.4062
$p(27)$:	0.0000	0.0000	0.5937	0.4062

Diskrétní Markovův řetězec a Markovův proces

Příklad

Sémantikou Markovova procesu je (nekonečná) množina všech trajektorií samplujících stavový prostor v (diskrétním) čase. Každá trajektorie poskytuje možnou dynamiku systému (s ohledem na pravděpodobnostní rozložení).



Markovův proces a stabilita

Mějme markovův proces charakterizovaný přechodovou maticí P. Rozložení p je *stabilní*, pokud platí

$$p = pP.$$

Stabilní rozložení je pevným bodem lineárního zobrazení určeného maticí P. Lze získat řešením homogení lineární soustavy:

$$\begin{aligned} p = pP &\Leftrightarrow p - pP = 0 \\ &\Leftrightarrow p(E - P) = 0 \end{aligned}$$

kde E je jednotková matice příslušného rozměru.

Markovův proces a biologické modely

- interpretace modelu
  - vycházíme z daného stavu (označkování)
  - v diskrétním časové kroku se projeví právě jedna (uschopněná) reakce r_i (s danou pravděpodobností) nebo stav zůstane nezměněn
  - každý přechod je vážen pravděpodobnostním rozložením závislým na konstantních pravděpodobnostech – při jeho respektování lze provádět Monte Carlo simulaci
- analyzační techniky
  - transientní analýza
    - výpočet distribuce $p(t)$ pro libovolné t
    - $p(t) = p(0)P^t$
  - stacionární analýza
    - analýza distribuce ve stabilním stavu

Stochastická sémantika modelu

- nedostatky Markovova procesu pro biologický model
  - jak nastavit pravděpodobnosti k hranám v $\mathit{reach}(\mathcal{N})$?
  - diskrétní samplování času neumožňuje dostatečně přesně vyjádřit rychlost reakce (četnost projevu reakce v čase)
- co chceme modelovat?
  - dynamika biologického systému při nízkých koncentracích
  - projev náhodných vlivů ovlivňujících provedení reakce
  - respektovat časo-prostorové jevy uvnitř buňky