Odhad Bellmanovy funkce pomocí sub-optimálních strategií

Jan Zeman

AS UTIA

7. března 2011

Dynamické programování

Systém - část světa, která je zajímavá pro rozhodovače

Rozhodovač - člověk nebo stroj s **cíli** zaměřenými na systém

Rozhodnutí u_t je navrženo rozhodovačem a aplikováno na systém

Výstup y_t pozorování systému dostupná rozhodovači

Účelová funkce g_t vyjadřuje míru dosažení cíle

Cíl najít způsob, jak zvolit $\{u_1, \ldots, u_T\}$, aby maximalizovala:

$$\max_{\{u_1,\ldots,u_T\}} E\left(\sum_{k=1}^{T} g_k\right).$$

Suboptimální strategie

Znalost v čase t:

$$\mathcal{P}_t = (u_1, \dots, u_{t-1}, y_1, \dots, y_t)$$

K návrhu rozhodnutí používáme maximální dostupnou znalost:

$$u_1(\mathcal{P}_1), u_2(\mathcal{P}_2), \dots, u_t(\mathcal{P}_t).$$

Představme posloupnost revizí minulých rozhodnutí:

$$\begin{array}{lrcl}
t = 1 & U_1(\mathcal{P}_1) & = & (u_1(\mathcal{P}_1)), \\
t = 2 & U_2(\mathcal{P}_2) & = & (u_1(\mathcal{P}_2), u_2(\mathcal{P}_2)), \\
t = 3 & U_3(\mathcal{P}_3) & = & (u_1(\mathcal{P}_3), u_2(\mathcal{P}_3), u_3(\mathcal{P}_3)), \\
 & & \vdots & \\
t = t & U_t(\mathcal{P}_t) & = & (u_1(\mathcal{P}_t), u_2(\mathcal{P}_t), u_3(\mathcal{P}_t), \dots, u_t(\mathcal{P}_t)).
\end{array}$$

tyto posloupnosti akcí nazveme *suboptimální strategie*.

Suboptimální strategie - pozorování

Čas	1	2	3	4	5	6	7	8	9	10
$U(\mathcal{P}_1)$	1									
$U(\mathcal{P}_2)$	-1	-1								
$U(\mathcal{P}_3)$	-1	-1	-1							
$U(\mathcal{P}_4)$	-1	-1	1	1						
$U(\mathcal{P}_5)$	-1	-1	0	-1	-1					
$U(\mathcal{P}_6)$	-1	-1	0	0	1	1				
$U(\mathcal{P}_7)$	-1	-1	0	0	1	1	1			
$U(\mathcal{P}_8)$	-1	-1	0	0	1	1	1	1		
$U(\mathcal{P}_9)$	-1	-1	0	0	1	1	1	-1	-1	
$U(\mathcal{P}_{10})$	-1	-1	0	0	1	1	1	1	-1	0
U^O	-1	-1	0	0	1	1	1	1	1	1

Suboptimální strategie se ustalují na optimálních rozhodnutích.
Nechť jsme schopni rozlišit, která část je optimální:

$$U(\mathcal{P}_t) = (\underbrace{u_1^O, u_2^O, \ldots, u_i^O}_{\text{optimální}}, u_{i+1}(\mathcal{P}_t), \ldots, u_t(\mathcal{P}_t))$$

Dynamické programování - Bellmanova funkce

Optimální řešení a Bellmanova rovnice:

$$\begin{aligned} u_t &= \arg\max_{u_t} \quad E\left(g_t + \mathcal{V}(\mathcal{P}_{t+1}) \quad |\mathcal{P}_t, u_t\right) \\ \mathcal{V}(\mathcal{P}_t) &= \max_{u_t} \quad E\left(g_t + \mathcal{V}(\mathcal{P}_{t+1}) \quad |\mathcal{P}_t, u_t\right) \end{aligned}$$

Pro výpočet přibližného řešení:

$$u_t = \arg\max_{u_t} \quad E\left(g_t|\mathcal{P}_t, u_t\right) + E(\mathcal{V}(\mathcal{P}_{t+1})|\mathcal{P}_t, u_t)$$

V následujícím postupu se zaměříme na hledání očekávané hodnoty Bellmanovy funkce:

$$E(\mathcal{V}(\mathcal{P}_{t+1})|\mathcal{P}_t, u_t)$$

Hledání Bellmanovy funkce

V čase t máme k dispozici soustavu Bellmanových rovnic:

$$\begin{aligned}
\mathcal{V}(\mathcal{P}_1) &= \max_{u_1} \quad E\left(g_1 + \mathcal{V}(\mathcal{P}_2) \quad |\mathcal{P}_1, u_1\right) \\
\mathcal{V}(\mathcal{P}_2) &= \max_{u_2} \quad E\left(g_2 + \mathcal{V}(\mathcal{P}_3) \quad |\mathcal{P}_2, u_2\right) \\
&\vdots \\
\mathcal{V}(\mathcal{P}_t) &= \max_{u_t} \quad E\left(g_t + \mathcal{V}(\mathcal{P}_{t+1}) \quad |\mathcal{P}_t, u_t\right)
\end{aligned}$$

Hledáme Bellmanovu funkci jako řešení této soustavy.

Nedostatky:

- Funkcionální rovnice
- Operace s maximem

Věřme, že umíme najít, že prvních i optimálních rozhodnutí:

$$
\begin{aligned}
\mathcal{V}(\mathcal{P}_1) &= \max_{u_1} \quad E\left(g_1 + \mathcal{V}(\mathcal{P}_2) \quad |\mathcal{P}_1, u_1\right) \\
\mathcal{V}(\mathcal{P}_2) &= \max_{u_2} \quad E\left(g_2 + \mathcal{V}(\mathcal{P}_3) \quad |\mathcal{P}_2, u_2\right) \\
&\vdots \\
\mathcal{V}(\mathcal{P}_i) &= \max_{u_i} \quad E\left(g_i + \mathcal{V}(\mathcal{P}_{i+1}) \quad |\mathcal{P}_i, u_i\right) \\
\mathcal{V}(\mathcal{P}_{i+1}) &= \max_{u_{i+1}} \quad E\left(g_{i+1} + \mathcal{V}(\mathcal{P}_{i+2}) \quad |\mathcal{P}_{i+1}, u_{i+1}\right) \\
&\vdots \\
\mathcal{V}(\mathcal{P}_t) &= \max_{u_t} \quad E\left(g_t + \mathcal{V}(\mathcal{P}_{t+1}) \quad |\mathcal{P}_t, u_t\right)
\end{aligned}
$$

Dosadíme je tedy do rovnic a navíc zohledníme, že spousta dat je již známá:

$$
\begin{aligned}
\mathcal{V}(\mathcal{P}_1) &= g_1 + \mathcal{V}(\mathcal{P}_2) \\
\mathcal{V}(\mathcal{P}_2) &= g_2 + \mathcal{V}(\mathcal{P}_3) \\
&\vdots \\
\mathcal{V}(\mathcal{P}_i) &= g_i + \mathcal{V}(\mathcal{P}_{i+1}) \\
\mathcal{V}(\mathcal{P}_{i+1}) &= \max_{u_{i+1}} \quad E\left(g_{i+1} + \mathcal{V}(\mathcal{P}_{i+2}) \quad |\mathcal{P}_{i+1}, u_{i+1}\right) \\
&\vdots \\
\mathcal{V}(\mathcal{P}_t) &= \max_{u_t} \quad E\left(g_t + \mathcal{V}(\mathcal{P}_{t+1}) \quad |\mathcal{P}_t, u_t\right)
\end{aligned}
$$

Suboptimální strategie a Bellmanova funkce

Získáme soustavu funkcionálních rovnic bez maximalizace:

$$\begin{aligned} \mathcal{V}(\mathcal{P}_1) &= g_1 + \mathcal{V}(\mathcal{P}_2) \\ \mathcal{V}(\mathcal{P}_2) &= g_2 + \mathcal{V}(\mathcal{P}_3) \\ &\vdots \\ \mathcal{V}(\mathcal{P}_i) &= g_i + \mathcal{V}(\mathcal{P}_{i+1}) \end{aligned}$$

Číslo i tedy charakterizuje počet rovnic a je žádoucí, aby rostl úměrně t.
Nyní můžeme zvolit parametrizovaného tvaru Bellmanovy funkce

$$\mathcal{V}_t = V_t(\Theta),$$

získáme soustavu rovnic pro neznámé parametry Θ.
Odhad Bellmanovy funkce je pak určen řešením této soustavy.

Volba správného počtu rovnic

Podařilo se ukázat dříve:

- Nejlepší je volit počet rovnic stejný, jako je počet optimálních akcí.
- Volit menší počet rovnic je maličko horší.
- Volit mnohem menší počet rovnic je o dost horší.
- Nejhorší je volit větší počet rovnic (funkce se počítá z neoptimálních hodnot).

Umělá data

Jak je vyrobit?

- Vezmeme reálná data
- Zvolíme typ AR modelu
- Naučíme se koeficienty a odhadneme disperzi šumu
- Na základě získaných koeficientů nagenerujeme nová data striktně se řídící AR modelem a použijeme více realizací šumu

Použitá reálná data:

- Kakao (CC)
- Lehká ropa (CL)
- Americký státní dluhopis (FV2)
- Japonský Yen (JY)
- Pšenice (W)

S čím se porovnáváme - boj s velkou nulou

Skládanka o dvou dílech:

$$u_t = \arg\max_{u_t} \quad E\left(\sum_{k=t}^{t+h} g_k | \mathcal{P}_t, u_t\right) + E(\mathcal{V}(\mathcal{P}_{t+h+1})|\mathcal{P}_t, u_t)$$

Volba ziskové funkce:

$$g_t = (y_t - y_{t-1})u_{t-1} + C|u_{t-1} - u_t|$$

Volba Bellmanovy funkce:

- Sub-optimální strategie (SOS)

$$\mathcal{V}(\mathcal{P}_t) := \Theta(u_t)(y_t, y_{t-1}, \ldots, y_{t-n}, 1)'$$

- Model predictive control (MPC) - tzv. velká nula

$$\mathcal{V}(\mathcal{P}_t) := 0$$

Přehled výsledků - Zeman vs. velká nula

Data generovaná z CC:

h	1	2	3	4	5	6	7	8	9	Celkově
SOS	15	13	6	8	3	5	3	3	2	0
Remíza	23	37	48	50	57	56	58	58	58	
MPC	27	15	11	7	5	4	4	4	5	1

Data generovaná z CL:

h	1	2	3	4	5	6	7	8	9	Celkově
SOS	25	23	23	16	16	17	15	14	15	0
Remíza	4	13	19	22	25	24	24	24	25	
MPC	36	29	23	27	24	24	26	27	25	1

Data generovaná z FV2:

h	1	2	3	4	5	6	7	8	9	Celkově
SOS	20	15	14	11	13	12	10	9	8	0
Remíza	33	36	36	42	40	43	44	47	47	
MPC	12	14	15	12	12	10	11	9	10	0

Přehled výsledků II. - Zeman vs. velká nula

Data generovaná z JY:

h	1	2	3	4	5	6	7	8	9	Celkově
SOS	12	11	4	1	0	0	0	0	0	1
Remíza	7	44	51	59	62	63	63	63	64	
MPC	46	10	10	5	3	2	2	2	1	0

Data generovaná z CC:

h	1	2	3	4	5	6	7	8	9	Celkově
SOS	14	10	3	1	1	0	1	1	1	0
Remíza	33	54	61	63	64	64	64	64	64	
MPC	18	1	1	1	0	1	0	0	0	0

Celkové skóre: 1:2 (vítězí MPC)

Závěrem

Práce do budoucna

- Zjistit, proč někde fungujeme a jinde ne (podezření na různé úrovně šumu)
- Projít myšlenkově algoritmus (Q-learning)
- Co nejpřesněji specifikovat, kdy lze sub-optimální strategie použít
- Co nejpřesněji specifikovat, jak hledat rozdělení na optimální a neoptimální akce