

Markovovy rozhodovací procesy

aneb jak se vypořádat s nejistotou

J. Větrovský¹ W. Bureš² J. Kuchařík³

¹SPŠE Ječná
Praha

²Polské gymnázium
Český Těšín

³Gymnázium Karla Čapka
Dobříš

Týden vědy na Jaderce, 2025

Obsah

1 Teorie

- Stochastický x deterministický
- Markovovy řetězce
- Markovovy rozhodovací procesy

2 Náš program

- Implementace
- Průchod stavovým prostorem

Obsah

1 Teorie

- Stochastický x deterministický
- Markovovy řetězce
- Markovovy rozhodovací procesy

2 Náš program

- Implementace
- Průchod stavovým prostorem

Stochastický x deterministický

- Deterministický stavový prostor
 - Hra dámy
- Stochastický stavový prostor
 - Opakované házení kostkou

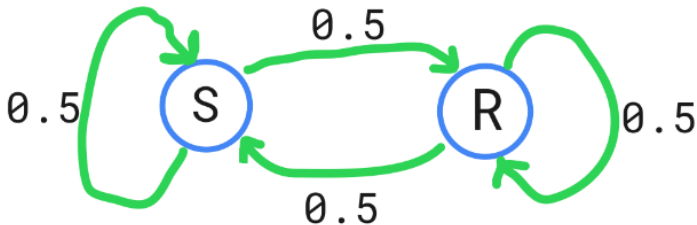
Stochastický x deterministický

- Deterministický stavový prostor
 - Hra dámy
- Stochastický stavový prostor
 - Opakované házení kostkou

Markovovy řetězce

- Typickým příkladem může být předpověď počasí na příští den

Markovovy řetězce



Markovovy rozhodovací procesy

Dovedou...

- modelovat vícedimenzionální stochastický stavový prostor
- propagovat závislost na předchozích rozhodnutích

Definice

Markovův rozhodovací proces je uspořádaná čtveřice $(S, A, P.(\cdot, \cdot), R.(\cdot, \cdot))$, kde:

- S je konečná množina stavů
- A je množina akcí v každém stavu
- $P_a(s, s')$ je pravděpodobnost, že se při akci a dostaneme ze stavu s do stavu s' , říkáme jí matice přechodu
- $R_a(s, s')$ je odměna, kterou dostaneme po přechodu do stavu s' ze stavu s

Algoritmus iterace hodnot

Postup

$$V(s) := R(s) + \gamma \max_a \sum_{s'} P_a(s, s') V(s'). \quad (1)$$

N-krát proveden pro matici $V(s)$.

Podle Bellmanova teorému[1] všechny tyto sumy konvergují.

Výstup

- Optimální strategie π
- Matice $V(s)$

Algoritmus iterace hodnot

Postup

$$V(s) := R(s) + \gamma \max_a \sum_{s'} P_a(s, s') V(s'). \quad (1)$$

N-krát proveden pro matici $V(s)$.

Podle Bellmanova teorému[1] všechny tyto sumy konvergují.

Výstup

- Optimální strategie π
- Matice $V(s)$

Obsah

1 Teorie

- Stochastický x deterministický
- Markovovy řetězce
- Markovovy rozhodovací procesy

2 Náš program

- Implementace
- Průchod stavovým prostorem

Implementace

Výpočet a Export

- C++
- Využití algoritmu iterace hodnot, výstupem π a $V(s)$
- Export do csv pro další využití

Visualizace

- Java
- Načtení csv a jednoduché okénkové zobrazení

Postup algoritmu iterace hodnot I

$\wedge$	$\wedge$	$>$	$+$
0	0	0	100
$\wedge$	$\#$	$<$	$-$
0		0	-100
$\wedge$	$\wedge$	$\wedge$	$\vee$
0	0	0	0

Postup algoritmu iterace hodnot II

$\wedge$	>	>	+
0	0	80	100
$\wedge$	#	$\wedge$	-
0		0	-100
$\wedge$	$\wedge$	$\wedge$	$\vee$
0	0	0	0

Postup algoritmu iterace hodnot III

>	>	>	+
0	64	88	100
^	#	^	-
0		54	-100
^	^	^	v
0	0	0	0

Postup algoritmu iterace hodnot IV

>	>	>	+
51.2	83.2	94.2	100
^	#	^	-
0		65.8	-100
^	>	^	<
0	0	43.2	0

Optimální strategie $R_a = -1$

>	>	>	+
94.9724	96.3787	97.6287	100
^	#	<	-
93.7224		88.658	-100
^	<	<	v
92.3162	91.0662	89.6875	79.6874

Optimální strategie $R_a = -3$

>	>	>	+
85.1819	89.4007	93.1507	100
^	#	^	-
81.4319		68.3562	-100
^	<	<	<
77.2132	73.4632	69.5624	47.3888

Optimální strategie $R_a = 0$

	^	^	+
100	100	100	100
^	#	<	-
100		100	-100
^	^	<	v
100	100	100	100

Optimální strategie $R_a = -300$

>	>	>	+
-1096.28	-674.408	-299.408	100
^	#	>	-
-1471.28		-494.67	-100
>	>	>	^
-1583.29	-1222.29	-847.288	-516.365

Optimální strategie $R_a = 50$

	^	<	+
499980	499980	499980	100
^	#	<	-
499980		499980	-100
^	^	^	v
499980	499980	499980	499980

Shrnutí

- Optimální strategii pohybu ve stochastickém stavovém prostoru jsme schopni nalézt pomocí Markovových rozhodovacích procesů
- Tuto strategii lze nalézt pomocí algoritmu iterace hodnot
- Můžeme tak pro daný stavový prostor získat strategii π
- Perspektiva
 - Ne vždy známe hodnoty R_a a matici přechodu P_a ,
 - Pro takové případy lze využít zpětnovazební učení

Děkujeme za pozornost

Zdroje I



Bellman, Richard

A Markovian decision process

Journal of Mathematics and Mechanics, pages 679–684