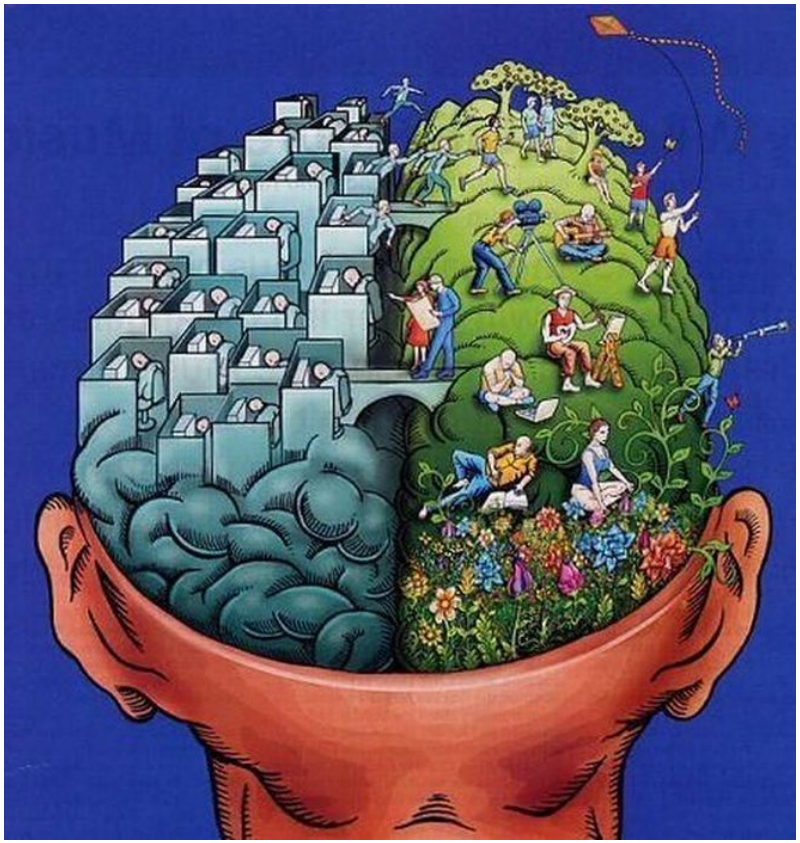
Mesterséges Intelligencia MI

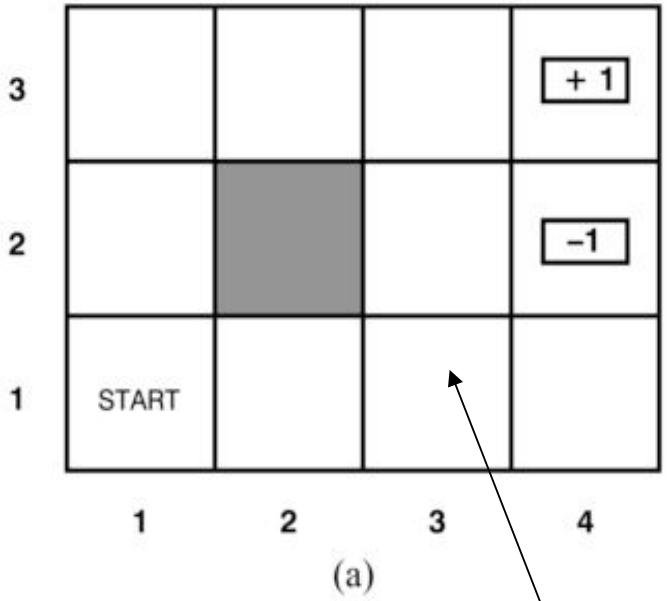
Komplex
döntések



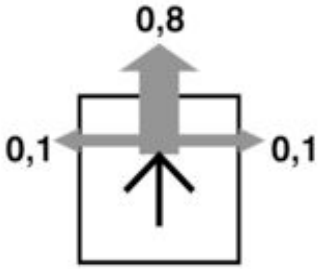
Dobrowiecki Tadeusz
Eredics Péter, és mások

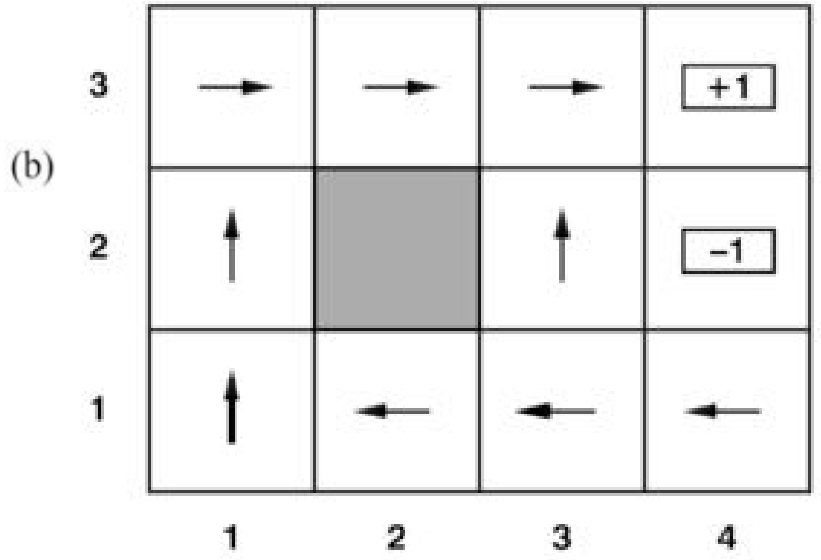
BME I.E. 437, 463-28-99
dobrowiecki@mit.bme.hu,
http://www.mit.bme.hu/general/staff/tade

Szekvenciális döntési probléma



- 0.04





http://mialmanach.mit.bme.hu/aima/ch17s01

Szekvenciális döntési probléma

Markov döntési folyamat

Kezdőállapot:	S0
Állapotátmenet-modell:	T(s, a, s′)
Jutalomfüggvény:	R(s), vagy R(s, a, s′)

Optimális eljárásmód = optimális mozgás, döntés cselekvés megválasztására, de nem elég egyszer, folyamatosan kell, amíg nincs a probléma vége.

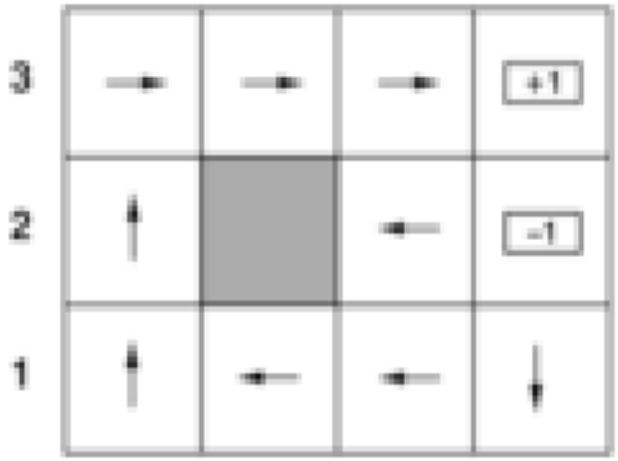
Szekvenciális döntési probléma

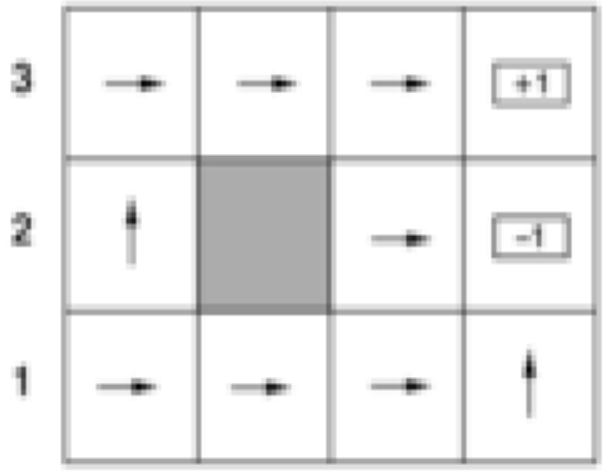
$R(s) \leq -1{,}6284$

$-0{,}4278 \leq R(s) \leq -0{,}0850$

$-0{,}0221 < R(s) < 0$

$R(s) > 0$



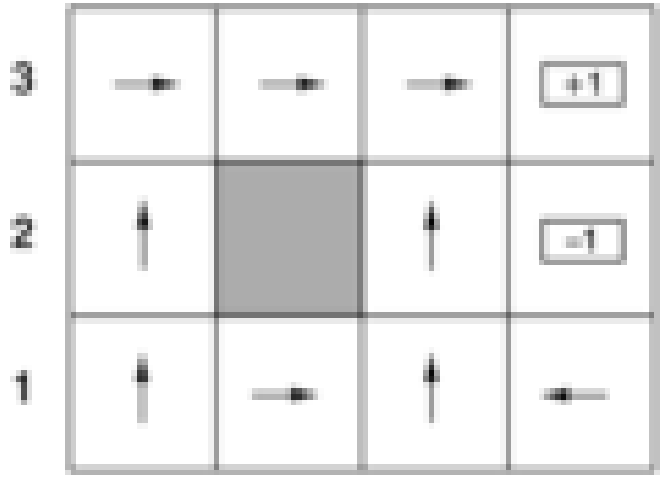


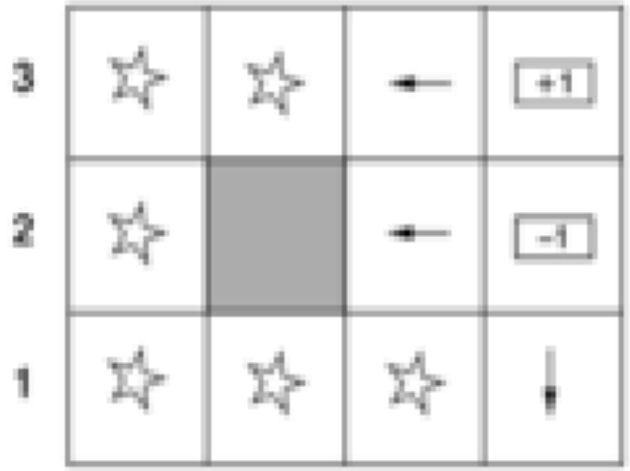
élet elviselhetetlen, ki!

élet kellemetlen; +1 állapot, –1 kockázattal

az élet csak kevéssé bánatos, ne legyen kockázat!

élet kifejezetten élvezhető, az ágens benn akar maradni





Szekvenciális döntési probléma

Optimalis szekvenciális döntési probléma

végtelen horizont

véges horizont

optimális eljárásmód nem-stacionárius

stacionárius

Többattribútumú hasznosságelmélet:

ágens preferenciái az állapotsorozatok között stacionáriusok

additív jutalmak $U_h([s_0, s_1, s_2, ...]) = R(s_0) + R(s_1) + R(s_2) + ...$

leszámítolt jutalmak

$$U_h([s_0, s_1, s_2, ...]) = R(s_0) + \gamma R(s_1) + \gamma^2 R(s_2) + ...$$

Szekvenciális döntési probléma

Leszámított jutalmak, egy végtelen sorozat hasznossága

$$U_h([s_0, s_1, s_2, \ldots]) = \Sigma_{t=0\ldots\infty} \gamma^t R(s_t) =< \Sigma_{t=0\ldots\infty} \gamma^t R_{\max} = {R_{\max}}/{(1-\gamma)}$$

Ha van végállapot, ha garantált, hogy az ágens végül bele kerül, akkor nincs szükség végtelen sorozatok összehasonlítására.
Egy eljárásmód, ami garantáltan végállapotba juttat, véges eljárásmód, $\gamma = 1$

Végtelen sorozatok összehasonlítása:
az időegységenkénti átlagjutalom

Optimális eljárásmód

$$\pi^* = \arg\max_{\pi} E[\Sigma_{t=0\ldots\infty} R(s_t) \mid \pi]$$

Szekvenciális döntési probléma

Értékiteráció

Egy **állapot hasznossága –** a belőle kiinduló állapotsorozatok várható hasznossága

Az állapotsorozatok függnek a végrehajtott eljárásmódtól, így elsőként egy adott π eljárásmódra definiáljuk a hasznosságot:

$$U^{\pi}(s) = E\left[\sum_{t=0}^{\infty} \gamma^t R(s_t) \,|\, \pi, s_0 = s\right]$$

$$\pi^*(s) = \arg\max_a \sum_{s'} T(s,a,s') U(s')$$

Optimális eljárásmód

Az **állapot hasznossága** - az állapotban tartózkodás közvetlen jutalmának és a következő állapot várható leszámított hasznosságának az összege, feltéve, hogy az ágens az optimális cselekvést választja (**Bellman** egyensúlyi **egyenlet**)

$$U(s) = R(s) + \gamma \max_a \sum_{s'} T(s,a,s') U(s')$$

Szekvenciális döntési probléma

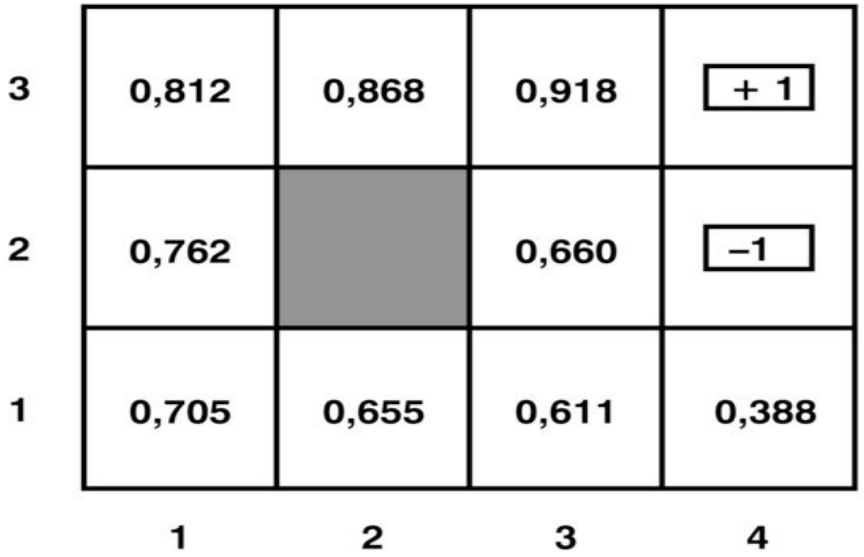
Legyen $\gamma = 1$ és a nem végállapotoknál $R(s) = -0{,}04$

Nézzük meg a 4 × 3-as világ Bellman-egyenleteinek egyikét.

Az (1, 1) állapothoz tartozó egyenlet:

$$U(1,1) = -0{,}04 + \gamma \max\{$$
$$0{,}8\,U(1,2) + 0{,}1\,U(2,1) + 0{,}1\,U(1,1) \quad \text{(Fel)},$$
$$0{,}8\,U(2,1) + 0{,}1\,U(1,2) + 0{,}1\,U(1,1) \quad \text{(Jobbra)},$$
$$0{,}9\,U(1,1) + 0{,}1\,U(1,2) \quad \text{(Balra)},$$
$$0{,}9\,U(1,1) + 0{,}1\,U(2,1)\} \quad \text{(Le)}$$

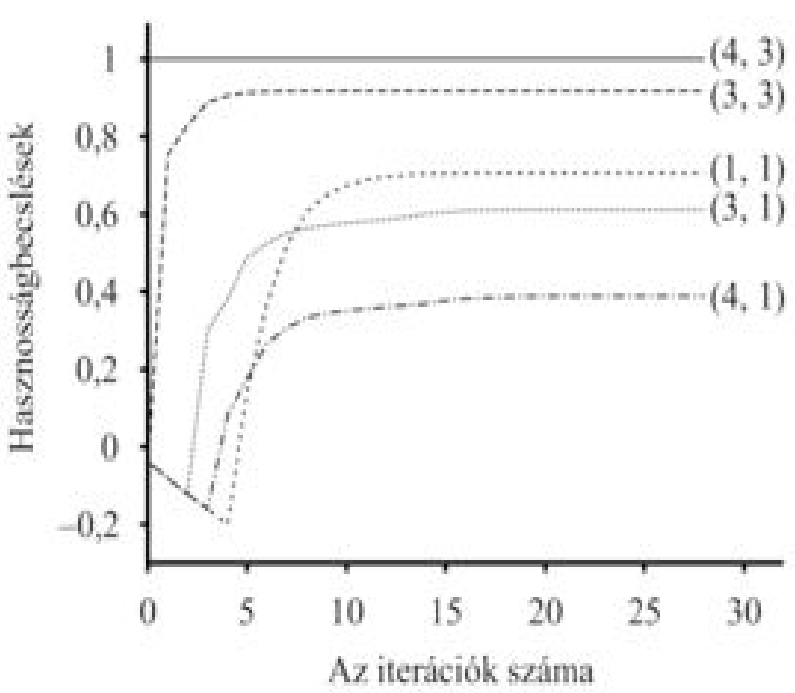
Sajnos nemlineáris – iteráció!



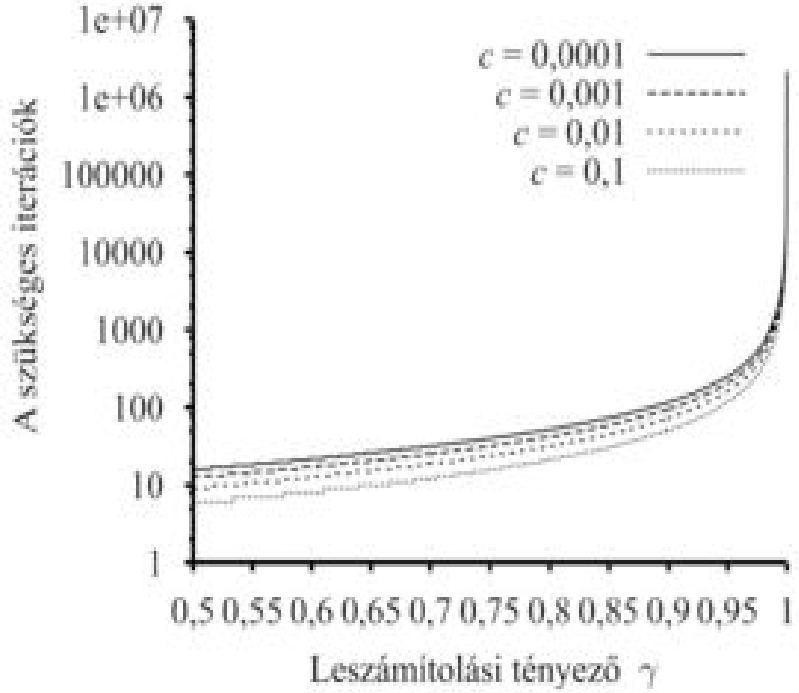
Szekvenciális döntési probléma

Bellman-frissítés:

$$U_{t+1}(s) = R(s) + \gamma \max_a \sum_{s'} T(s,a,s') U_t(s')$$



A hasznosságok fejlődése



A szükséges értékiterációk száma, hogy a hiba garantáltan legfeljebb $\varepsilon = c\, R_{max}$ legyen

Szekvenciális döntési probléma

Az értékiteráció konvergenciája – **kontrakció** $U_{i+1} = B\,U_i$

$$\|U\| = \max_s |U(s)|$$

$$\|BU_i - U^*\| \le \gamma \|U_i - U^*\|, \qquad U^* \text{ az igazi: } B(U^*) = U^*$$

Az összes állapot hasznossága korlátos $\pm R_{max}/(1-\gamma)$ értékkel A maximális kezdeti hiba $\|U_0 - U^*\| \le 2R_{max}/(1-\gamma)$

Ha $\|U_{i+1} - U_i\| < \varepsilon(1-\gamma)/\gamma$, akkor $\|U_{i+1} - U^*\| < \varepsilon$

Szekvenciális döntési probléma

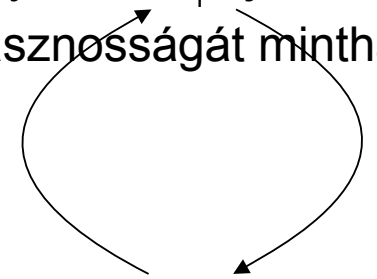
Eljárásmód-iteráció

(optimális eljárásmódot is kaphatunk, ha a hasznosságfüggvény becslése pontatlan - ha egy cselekvés egyértelműen jobb, mint a többi, akkor a releváns állapotok pontos hasznosságát nem szükséges precízen tudnunk)

$$U(s) = R(s) + \gamma \max_a \sum_{s'} T(s,a,s')U(s')$$

Eljárásmód-értékelés

Egy adott π_i eljárásmódnál számítsuk ki $U_i = U^{\pi i}$-t, az egyes állapotok hasznosságát mintha π_i volna végrehajtva.



lineáris!

$$U_t(s) = R(s) + \gamma \sum_{s'} T(s,\pi_t(s),s')U_t(s')$$

$$U_{t+1}(s) \leftarrow R(s) + \gamma \sum_{s'} T(s,\pi_t(s),s')U_t(s')$$

Eljárásmód-javítás

Módosított eljárásmód-iteráció

for each s állapotra **in** S **do**

if $\max_a \sum_{s'} T(s,a,s')U[s'] > \sum_{s'} T(s,\pi[s],s')U[s']$ **then**

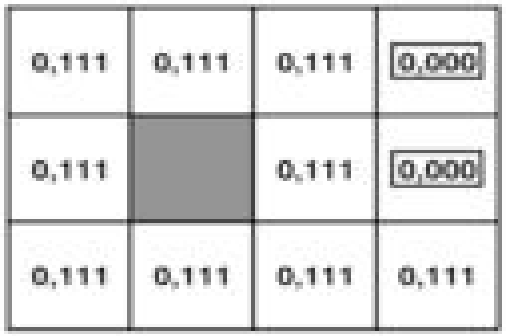
$$\pi[s] \leftarrow \text{argmax}_a \sum_{s'} T(s,a,s')U[s']$$

Szekvenciális döntési probléma

Részlegesen megfigyelhető Markov döntési folyamat

Kezdőállapot: S0
Állapotátmenet-modell: T(s, a, s′)
Jutalomfüggvény: R(s), v. R(s, a, s′)
Megfigyelési modell,
az s állapotban az o megfigyelés érzékelésének a valószínűsége
O(s, o)

Hiedelmi állapot = b(s) = eloszlás állapotok felett



$$\left\langle \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, 0, 0 \right\rangle$$

$$b'(s') = \alpha O(s', o) \sum_{s} T(s, a, s') b(s)$$

(szűrés)

Szekvenciális döntési probléma

$$P(o \mid a,b) = \sum_{s'} P(o \mid a,s',b)\, P(s' \mid a,b)$$

$$= \sum_{s'} O(s',o) P(s' \mid a,b) = \sum_{s'} O(s',o) \sum_{s} T(s,a,s') b(s)$$

$$\tau(b,a,b') = P(b' \mid a,b) = \sum_{o} P(b' \mid o,a,b)\, P(o \mid a,b)$$

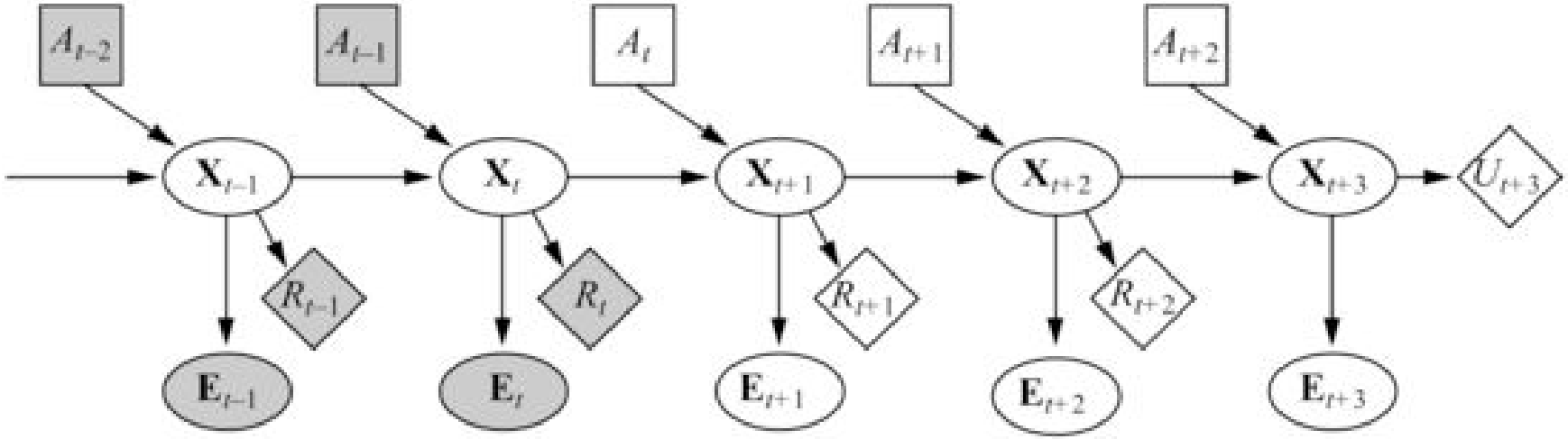
$$= \sum_{o} P(b' \mid o,a,b) \sum_{s'} O(s',o) \sum_{s} T(s,a,s') b(s)$$

$$\rho(b) = \sum_{s} b(s) R(s)$$

RMMDF megoldása a fizikai (véges) állapottérben redukálható egy MDF megoldására a hozzá tartozó hiedelmi állapot térben (val. eloszlások folytotonos terében)

Szekvenciális döntési probléma

Dinamikus Döntési Hálók (nem foglalkozunk vele)



Játékelmélet

(vele sem)

	Aliz:tanúskodik	*Aliz:tagad*
Bendegúz:tanúskodik	$A = -5, B = -5$	$A = -10, B = 0$
Bendegúz:tagad	$A = 0, B = -10$	$A = -1, B = -1$

Működési mód tervezés (vele sem)

(egy olyan játék tervezése, aminek a megoldása az egyes ágensek által követett saját racionális stratégiáik együttese, és ez egy globális hasznosságfüggvény maximálását eredményezi, pl. árverés)