Tik-79.230 **Kevät 2001**
Agenttipohjaisen tietojenkäsittelyn perusteet
Laskuharjoitus 4
Ratkaisut

1. Olkoon agentin tila-avaruus S ja sen mahdollisten toimintojen joukko A. Tällöin *politiikka* P on kuvaus $S \to A$. Toisin sanoen, politiikka liittää jokaiseen agentin tilaan yksikäsitteisen toiminnon, joka suoritetaan aina kun agentti on kyseisessä tilassa.

   Politiikka P on *optimipolitiikka*, mikäli kuhunkin tilaan liitetty toiminto maksimoi agentin hyötyfunktion siinä tilassa.

   Tehtävässä $S = \{(1,1),(2,1),(3,1),(2,2),(3,2)\}$ ja $A = \{\leftarrow, \uparrow, \rightarrow, \downarrow\}$.

		+1
S		-1

   (a) Arvoiteraatiossa lähdetään laskemaan jokaisen tilan jokaisen toiminnon hyötyä. Toiminnon hyöty saadaan laskemalla odotusarvo niiden tilojen hyötyarvoista, joihin on mahdollista päästä kyseisellä toiminnolla.

   Kunkin tilan politiikaksi valitaan toiminto, jonka hyötyfunktion odotusarvon on suurin, ja tilan hyötyarvoksi asetetaan

$$U(i) = R(i) + \max_a \sum_j M_{ij}^a U(j),$$

   missä $R(i)$ on tilan *palkinto* (tässä 1 tilassa $(3,2)$, -1 tilassa $(3,1)$ ja -0.2 kaikissa muissa tiloissa) ja $\max_a \sum_j M_{ij}^a U(j)$ on valitun toiminnon hyödyn odotusarvo.

   Tämän jälkeen toistetaan arvojen laskemista niin kauan, että tilojen hyötyarvot lopulta vakiintuvat.

   Kierros 1.

Tila	Toiminto	Odotusarvot	
$(2,2)$	$\leftarrow$	$1 \cdot (-0.2) = -0.2$	
	$\uparrow$	$0.9 \cdot (-0.2) + 0.1 \cdot 1 = -0.08$	
	$\rightarrow$	$0.8 \cdot 1 + 0.2 \cdot (-0.2) = 0.76$	$\times$
	$\downarrow$	$0.9 \cdot (-0.2) + 0.1 \cdot 1 = -0.08$	
$(2,1)$	$\leftarrow$	$1 \cdot (-0.2) = -0.2$	$\times$
	$\uparrow$	$0.9 \cdot (-0.2) + 0.1 \cdot (-1) = -0.28$	
	$\rightarrow$	$0.8 \cdot (-1) + 0.2 \cdot (-0.2) = -0.84$	
	$\downarrow$	$0.9 \cdot (-0.2) + 0.1 \cdot (-1) = -0.28$	
$(1,1)$	$\leftarrow$	$1 \cdot (-0.2) = -0.2$	
	$\uparrow$	$1 \cdot (-0.2) = -0.2$	
	$\rightarrow$	$1 \cdot (-0.2) = -0.2$	
	$\downarrow$	$1 \cdot (-0.2) = -0.2$	

Tilan (2,2) politiikaksi saatiin $\rightarrow$ ja tilan (2,1) politiikaksi saatiin $\leftarrow$. Koska kaikilla tilan (1,1) toiminnoilla on tismalleen sama oletettu hyötyarvo, voi tilan politiikan valita mielivaltaisesti.

■	$\rightarrow$	+1
S	$\leftarrow$	-1

Tilojen uusiksi hyötyarvoiksi saadaan:

$$U(2,2) = -0.2 + 0.76 = 0.56$$
$$U(2,1) = -0.2 - 0.2 = -0.4$$
$$U(1,1) = -0.2 - 0.2 = -0.4$$

Kierros 2.

Tila	Toiminto	Odotusarvot	
$(2,2)$	$\leftarrow$	$0.9 \cdot 0.56 + 0.1 \cdot (-0.4) = 0.464$	
	$\uparrow$	$0.9 \cdot 0.56 + 0.1 \cdot 1 = 0.604$	
	$\rightarrow$	$0.8 \cdot 1 + 0.1 \cdot 0.56 + 0.1 \cdot (-0.4) = 0.816$	$\times$
	$\downarrow$	$0.8 \cdot (-0.4) + 0.1 \cdot 0.56 + 0.1 \cdot 1 = -0.164$	
$(2,1)$	$\leftarrow$	$0.9 \cdot (-0.4) + 0.1 \cdot 0.56 = -0.304$	
	$\uparrow$	$0.8 \cdot 0.56 + 0.1 \cdot (-1) + 0.1 \cdot (-0.4) = 0.308$	$\times$
	$\rightarrow$	$0.8 \cdot (-1) + 0.1 \cdot (-0.4) + 0.1 \cdot 0.56 = -0.784$	
	$\downarrow$	$0.9 \cdot (-0.4) + 0.1 \cdot (-1) = -0.46$	
$(1,1)$	$\leftarrow$	$1 \cdot (-0.4) = -0.4$	
	$\uparrow$	$1 \cdot (-0.4) = -0.4$	
	$\rightarrow$	$1 \cdot (-0.4) = -0.4$	
	$\downarrow$	$1 \cdot (-0.4) = -0.4$	

Uudeksi politiikaksi saadaan:

■	$\rightarrow$	+1
S	$\uparrow$	-1

ja uusiksi hyötyarvoiksi

$$U(2,2) = -0.2 + 0.816 = 0.616$$
$$U(2,1) = -0.2 + 0.308 = 0.108$$
$$U(1,1) = -0.2 - 0.4 = -0.6$$

Jatkettaessa samaan tapaan pysyvät seuraavalla kierroksella tilojen (2,2) ja (2,1) politiikat ennallaan. Tilalle (1,1) saadaan vihdoinkin politiikka, sillä nyt tilan (2,1) hyötyarvo on korkeampi kuin tilan (1,1). Näinollen alkutilasta kannattaa lähteä oikealle.

■	$\rightarrow$	+1
$\rightarrow$	$\uparrow$	-1

Tämän jälkeen politiikka ei enää muutu, mutta arvojen vakiintumiseen tarvitaan vielä huomattavan monta laskentakierrosta.

(b) Politiikkaiteraatiota käytettäessä arvotaan ensin tiloille satunnainen politiikka. Tämän jälkeen lasketaan, mitkä ovat tilojen hyötyarvot tällä politiikalla ja päivitetään politiikkaa siten, että valitaan se toiminto, jonka odotusarvo on aikaisemmin laskettujen arvojen mukaan paras. Näin jatketaan, kunnes politiikka ei enää muutu.

Valitaan aloituspolitiikka satunnaisesti seuraavanlaiseksi:

	↑	+1
→	←	-1

Lasketaan ylläolevan politiikan mukaiset hyötyarvot tiloille ratkaisemalla seuraava yhtälöryhmä. (Tässä u_{ij} tarkoittaa tilan (i, j) hyötyarvoa).

$$\begin{aligned} u_{11} &= 0.2u_{11} + 0.8u_{21} - 0.2 \\ u_{21} &= 0.8u_{11} + 0.1u_{21} + 0.1u_{22} - 0.2 \\ u_{22} &= 0.9u_{22} + 0.1 \cdot 1 - 0.2 \end{aligned}$$

Yhtälöryhmän ratkaisuksi saadaan:

$$\begin{aligned} u_{11} &= -5.25 \\ u_{21} &= -5 \\ u_{22} &= -1 \end{aligned}$$

Lasketaan ylläolevien hyötyarvojen perusteella odotusarvot eri toiminnoille kussakin tilassa:

Tila	Toiminto	Odotusarvot	
$(2,2)$	$\leftarrow$	$0.9 \cdot (-1) + 0.1 \cdot (-5) = -1.4$	
	$\uparrow$	$0.9 \cdot (-1) + 0.1 \cdot 1 = -0.8$	
	$\rightarrow$	$0.8 \cdot 1 + 0.1 \cdot (-1) + 0.1 \cdot (-5) = 0.2$	$\times$
	$\downarrow$	$0.8 \cdot (-5) + 0.1 \cdot (-1) + 0.1 \cdot 1 = -4$	
$(2,1)$	$\leftarrow$	$0.9 \cdot (-5.25) + 0.1 \cdot (-1) = -4.825$	
	$\uparrow$	$0.8 \cdot (-1) + 0.1 \cdot (-1) + 0.1 \cdot (-5.25) = -1.425$	
	$\rightarrow$	$0.8 \cdot (-1) + 0.1 \cdot (-5) + 0.1 \cdot (-1) = -1.4$	$\times$
	$\downarrow$	$0.8 \cdot (-5) + 0.1 \cdot (-1) + 0.1 \cdot (-5.25) = -4.625$	
$(1,1)$	$\leftarrow$	$1 \cdot (-5.25) = -5.25$	
	$\uparrow$	$0.9 \cdot (-5.25) + 0.1 \cdot (-5) = -5.225$	
	$\rightarrow$	$0.8 \cdot (-5) + 0.2 \cdot (-5.25) = -5.05$	$\times$
	$\downarrow$	$0.9 \cdot (-5.25) + 0.1 \cdot (-5) = -5.225$	

Uudeksi politiikaksi saadaan:

	→	+1
→	→	-1

Seuraavalla iteraatiokerralla saadaan tilan $(2, 1)$ politiikka käännettyä optimaaliseen arvoon $\uparrow$.

2. Kun tutkitaan agentin uskomusten kehittymistä ajan kuluessa, jaetaan jokaisen ajanhetken käsittely tavallisesti kahteen osaan:

   - *Ennustusvaihe*, jolloin arvioidaan tämänhetkisten uskomusten perusteella ($Bel(\mathbf{X_t})$), millainen maailma tulee olemaan seuraavalla hetkellä $t+1$. Näin saatua todennäköisyysjakaumaa merkitään $\widehat{Bel}(\mathbf{X_{t+1}})$.

   $$\widehat{Bel}(\mathbf{X_{t+1}}) = \sum_{\mathbf{X_t}} \mathbf{P}(\mathbf{X_{t+1}} \mid \mathbf{X_t} = \mathbf{x_t}) Bel(\mathbf{X_t} = \mathbf{x_t})$$

   - *Arviointivaihe*, jolloin saadaan hetken $t+1$ havainnot $\mathbf{E_{t+1}}$. Ne liitetään edellisessä vaiheessa laskettuun ennustukseen käyttäen Bayesin päivitystä:

   $$Bel(\mathbf{X_{t+1}}) = \alpha \mathbf{P}(\mathbf{E_{t+1}} \mid \mathbf{X_{t+1}}) \widehat{Bel}(\mathbf{X_{t+1}}).$$

   Tässä α on normalisointivakio, jolla kertomalla saadaan todennäköisyyksien summa ykköseksi.

Kirjoitetaan tehtävanannossa annetut todennäköisyysjakaumat selkeämpään muotoon:

	U_{t+1}	$\neg U_{t+1}$
U_t	0.6	0.4
$\neg U_t$	0.1	0.9

	V_{t+1}	$\neg V_{t+1}$
V_t	0.6	0.4
$\neg V_t$	0.3	0.7

Ennustetaan aluksi, millainen tilanne on hetkellä t_1, kun hetken t_0 todennäköisyysjakaumat tunnetaan ($Bel(U_0) = 0.1, Bel(V_0) = 0.3$).

$$\begin{aligned}
\widehat{Bel}(U_1) &= P(U_1 \mid \neg U_0)Bel(\neg U_0) + P(U_1 \mid U_0)Bel(U_0) \\
&= 0.9 \cdot 0.1 + 0.1 \cdot 0.6 = 0.15 \\
\widehat{Bel}(\neg U_1) &= P(\neg U_1 \mid \neg U_0)Bel(\neg U_0) + P(\neg U_1 \mid U_0)Bel(U_0) \\
&= 0.85 \\
\widehat{Bel}(V_1) &= P(V_1 \mid \neg V_0)Bel(\neg V_0) + P(V_1 \mid V_0)Bel(V_0) \\
&= 0.39 \\
\widehat{Bel}(\neg V_1) &= P(\neg V_1 \mid \neg V_0)Bel(\neg V_0) + P(\neg V_1 \mid V_0)Bel(V_0) \\
&= 0.61
\end{aligned}$$

Hetkellä t_1 havaitaan, että molemmat ovet ovat auki, eli havainnoiksi saadaan $A_1 \wedge B_1$. Seuraavaksi lasketaan, kuinka todennäköistä on, että ovi U on todellakin auki hetkellä t_1, kun havainnoitsija on näkevinään sen olevan auki.

$$\begin{aligned}
Bel(U_1) &= \alpha_{U_1} P(A_1 \mid U_1)\widehat{Bel}(U_1) \\
&= \alpha_{U_1} \cdot 0.9 \cdot 0.15 = 0.135\alpha_{U_1} \\
Bel(\neg U_1) &= \alpha_{U_1} P(A_1 \mid \neg U_1)\widehat{Bel}(\neg U_1) \\
&= \alpha_{U_1} \cdot 0.1 \cdot 0.85 = 0.085\alpha_{U_1}
\end{aligned}$$

Lasketaan normalisointivakion α_{U_1} suuruus

$$\alpha_{U_1} = \frac{1}{0.135 + 0.085} = 4.54$$

Tästä saadaan

$$\begin{aligned} Bel(U_1) &= 0.61 \\ Bel(\neg U_1) &= 0.39 \end{aligned}$$

Lasketaan samat arvot myös ovelle V:

$$\begin{aligned} Bel(V_1) &= \alpha_{V_1} P(B_1 \mid V_1)\widehat{Bel}(V_1) \\ &= 0.27 \cdot \alpha_{V_1} \\ Bel(\neg V_1) &= \alpha_{V_1} P(B_1 \mid \neg V_1)\widehat{Bel}(\neg V_1) \\ &= 0.06 \cdot \alpha_{V_1} \\ \alpha_{V_1} &= \frac{1}{0.27 + 0.06} = 3.03 \\ Bel(V_1) &= 0.82 \\ Bel(\neg V_1) &= 0.18 \end{aligned}$$

Nyt on siis saatu laskettua, kuinka todennäköisesti ovet ovat auki havainnoitsijan mielestä hetkellä t_1. Nyt voidaan käyttää täysin samaa proseduuria selvittämään uskomukset hetkellä t_2, kun molemmat ovet ovat havainnoitsijan mielestä kiinni ($\neg A_2 \wedge \neg B_2$).

$$
\begin{aligned}
\widehat{Bel}(U_2) &= P(U_2 \mid U_1)Bel(U_1) + P(U_2 \mid \neg U_1)Bel(\neg U_1) \\
&= 0.6 \cdot 0.61 + 0.39 \cdot 0.1 = 0.41 \\
\widehat{Bel}(\neg U_2) &= 0.59 \\
Bel(U_2) &= \alpha_{U_2} P(\neg A_2 \mid U_2)\widehat{Bel}(U_2) \\
&= 0.04\alpha_{U_2} \\
Bel(\neg U_2) &= \alpha_{U_2} P(\neg A_2 \mid \neg U_2)\widehat{Bel}(\neg U_2) \\
&= 0.53\alpha_{U_2} \\
\alpha_{U_2} &= \frac{1}{0.04 + 0.53} \\
Bel(U_2) &= 0.07 \\
Bel(\neg U_2) &= 0.93 \\
\widehat{Bel}(V_2) &= P(V_2 \mid V_1)Bel(V_1) + P(V_2 \mid \neg V_1)Bel(\neg V_1) \\
&= 0.55 \\
\widehat{Bel}(\neg V_2) &= 0.45 \\
Bel(V_2) &= \alpha_{V_2} P(\neg A_2 \mid V_2)\widehat{Bel}(V_2) \\
&= 0.16\alpha_{V_2} \\
Bel(\neg V_2) &= \alpha_{V_2} P(\neg A_2 \mid \neg V_2)\widehat{Bel}(\neg V_2) \\
&= 0.41\alpha_{V_2} \\
\alpha_{V_2} &= \frac{1}{0.16 + 0.41} \\
Bel(V_2) &= 0.28 \\
Bel(\neg V_2) &= 0.72
\end{aligned}
$$

Tehtävässä kysyttiin, kuinka varma havannoitsija voi olla, että hetkellä t_2 ainakin toinen ovista on kiinni.

$$
\begin{aligned}
Bel(\neg(U_2 \wedge V_2)) &= 1 - Bel(V_2)Bel(U_2) \\
&= 1 - 0.01 = 0.99
\end{aligned}
$$