

- * Hyödyn aksioomat eivät määrää yksikäsitteistä hyötyfunktiota
- * Voimme esim. tehdä tiolle $U(S)$ lineaarisen muunnoksen

$$U'(S) = k_1 + k_2 U(S)$$

$(k_1$ on vakio, k_2 on mv. positiivinen vakio) ilman, että agentin käyttäytyminen muuttuu

- * Deterministisessä maailmassa, jossa ei ole arvonvoja, mikä tahansa monotoinen muunnos säilyttää agentin käytöksen

- * Esim. $\sqrt[3]{U'(S)}$

- * Hyötyfunktio on tällöin ordinaalinen — se antaa tiloilte järjestyksen, numeerisilla arvoilla ei ole merkitystä

228

- * Hyötyarvojen skaala käy parhaasta mahdollisesta palkinnosta u_T pähmipaan katastoihin u_L
- * *Normalisoidulla hyödyllä* $u_L = 0$ ja $u_T = 1$

- * Ääniarvojen väliin jäävän tilan S arvotamiseksi agentti voi verrata sitä standardiarvoon $[p, u_T; 1 - p, u_L]$
- * Todennäköisyyttä p on säädettävä kunnes kunnnes agentin mielestä standardiarvonta ja S ovat samanarvoisia
- * Jos käytössä on normalisoidut hyödyt, niin lopullinen p on S :n hyötyarvo
- * Usein hyötyarvo on monen muuttujan (attribuuttien) $\vec{x} = (x_1, \dots, x_n)$ arvojen $\vec{x} = (x_1, \dots, x_n)$ määräämä

229

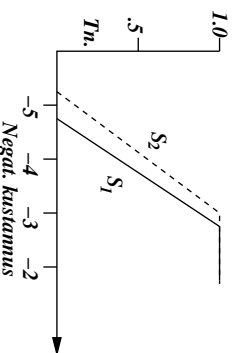
- * Tarkastelemaan tilannetta, missä muuden arvojen ollessa samat, attribuuttien korkeampi arvo tietää myös korkeampaa hyötyfunktion arvoa
- * Jos attribuuttivektoreille \vec{x} ja \vec{y} pätee $x_i \geq y_i \forall i$, niin \vec{x} *dominoi* (*aldissti*) \vec{y} :tä
- * Jos esim. lentokentän mahdollinen sijoituspaikka S_1 on halvempi, tuottaa vähemmän äänisaastetta ja on turvallisempi kuin S_2 , niin jälkimmäistä ei enää tarvitse harkita
- * Epävarmuuden vallitessa aidot domivoitusuhteet ovat harvinaisempia kuin deterministisessä tapauksessa
- * *Stokastinen dominanssi* on usein käytökelpoinen vertailutapa

230

- * Jos lentokentän sijoittamiskustannuksen uskotaan olevan tasaisesti jakautunut välille

- ◊ S_1 : 2.8 ja 4.8 miljoonaa euroa
- ◊ S_2 : 3.0 ja 5.2 miljoonaa euroa

niin kumulatiivisia jakaumia tarkastelemalla havaitaan S_1 :n dominoivan stokastisesti S_2 :ta (koska kustannukset ovat negatiivisia)



231

- * Kumulatiivinen jakauma on alkuperäisen jakauman integraali
 - * Olk. tapahtumien A_1 ja A_2 jakaumat attribuutille X $p_1(x)$ ja $p_2(x)$
 - * A_1 dominoi stokastisesti A_2 :ta, jos
- $$\forall x \int_{-\infty}^x p_1(x') dx' \leq \int_{-\infty}^x p_2(x') dx'$$
- * Jos
 - ◊ A_1 dominoi stokastisesti A_2 :ta ja $U(x)$ on mv. monotonisesti vähenevä hyötyfunktio,
 - ◊ niin A_1 :n odotusarvoinen hyöty on väh. yhtä korkea kuin A_2 :n

232

- * Jos jokin toiminto on toisen dominoiva kaikkien attribuuttien suhteen, niin se voidaan jättää huomiotta

Information arvo

- * Öljykenttään myydään porausoikeuksia, palstoja on n kappaletta, mutta vain yhdessä niistä on C euron edestä öljyä
- * Yhden palstan hinta on C/n euroa
- * Seismologi tarjoaa yritykselle tutkimustietoa palstasta nro 3, joka paljastaa aukottomasti onko palstalla öljyä vai ei
- * Paljonko yrityksen kannattaa maksaa tiedosta?
- * Th.:llä $1/n$ tutkimus kertoo palstalla 3 olevan öljyä, jolloin yritys hankkii sen hintaan C/n ja ansaitsee $(n-1)C/n$ euroa
- * Th.:llä $(n-1)/n$ tutkimus osoittaa, ettei palsta 3 sisällä öljyä, jolloin yhtiö hankkii jonkin muista palstoista

233

- * Koska palstan 3 tilanne jo tunnetaan, niin ostetulla palstalla löytyy nyt öljyä th.:llä $1/(n-1)$, joten yhtiön odotusarvoinen voitto on $C/(n-1) - C/n = C/n(n-1)$ euroa

Odotusarvoinen voitto annettuna tutkimustulos on siis

$$\frac{1(n-1)C}{n} + \frac{n-1}{n} \frac{C}{n(n-1)} = \frac{C}{n}$$

- * Seismologille siis kannattaa maksaa aina palstan hintaan asti
- * Lisäinformaatio on arvokasta, koska sen avulla toiminta voidaan sopeuttaa valitsevaan tilanteeseen
- * Ilman informaatiota on trydyttävä kaikissa mahdollisissa tilanteissa keskimäärin parhaaseen toimintaan

234

- * Olk. E_j on satunnaisuuttuja, jonka arvosta saadaan uusi tarkka havainto
- * Agentin aiempi itätämys on E
- * Ilman lisäinformaatiota parhaan toiminnon α arvo on $EU(\alpha | E) =$

$$\max_{\alpha} \sum_i U(\text{Tulos}_i(A)) \\ P(\text{Tulos}_i(A) | \text{Suoritus}(A), E)$$

- * Uusi havainto muuttaa parhaan toiminnon ja sen arvon
- * Mutta toistaiseksi E_j on satunnaisuuttuja, jonka arvoa ei tunneta, joten voimme vain summata yli sen kaikkien mahdollisten arvojen e_{jk}
- * Havainnon E_j arvo on lopulta

$$\left(\sum_k P(E_j = e_{jk} | E) EU(\alpha_{e_{jk}} | E, E_j = e_{jk}) \right) - EU(\alpha | E)$$

235

$$-0.0221 < R(s) < 0$$

| | | |
|---|---|----|
| → | → | +1 |
| ↑ | ■ | -1 |
| → | → | ↑ |

$$-0.4278 < R(s) < -0.0850$$

| | | |
|---|---|----|
| → | → | +1 |
| ↑ | ■ | -1 |
| → | → | ↑ |

244

- * Äärettömän horisontin tapauksessa agentin toiminta-ajalle ei ole asetettu ylärajaa
- * Jos toiminta-aika on rajoitettu, niin eri aikoina samassa tilassa voidaan joutua tekemään eri toimintopäätöksiä — optimaalinen politiikka ei ole *stationäärinen*

- * Sen sijaan äärettömän horisontin tapauksessa ei ole syytä muuttaa tilan toimintaa kerrasta toiseen, joten optimaalinen politiikka on stationäärinen
- * Tilajonon s_0, s_1, s_2, \dots *diskontattu* palkkio on

$$R(s_0) + \gamma R(s_1) + \gamma^2 R(s_2) + \dots,$$

missä $0 \leq \gamma \leq 1$ on diskonttaustekijä

245

- * Kun $\gamma = 1$, niin ympäristöhistorian palkkioksi saadaan erikoistapauksena additiivinen palkkio

- * Lähellä nolaa oleva γ puolestaan tarkoittaa tulevien palkkioiden merkityksen pienenemistä

- * Jos äärettömän horisontin maali-massa ei ole lainkaan saavutettava maallitilaa, niin ympäristöhistorioista tulee äärettömän pitkiä

- * Additiivisilla palkkioilla myös hyödyt kasvavat yleisesti ottaen äärettömäksi

- * Diskontatulla palkkioilla ($\gamma < 1$) äärettömänkin jonon palkkio on äärellinen

246

- * Olk. R_{\max} palkkioiden ylärajaa. Tällöin geometrisen sarjan summana

$$\sum_{t=0}^{\infty} \gamma^t R(s_t) \leq \sum_{t=0}^{\infty} \gamma^t R_{\max} = R_{\max} / (1 - \gamma)$$

- * Keelvoiminen politiikka (proper policy) takaa agentin pääsevän lopputilaan silloin kun ympäristössä on lopputiloja
- * Tällöin äärettömistä tilajonoista ei ole huolta ja voidaan jopa käyttää additiivisia palkkioita

- * Optimaalinen politiikka diskontatulla palkkioilla on

$$\pi^* = \arg \max_{\pi} E \left[\sum_{t=0}^{\infty} \gamma^t R(s_t) \mid \pi \right],$$

missä odotusarvo lasketaan yli kaikkien mahdollisten tilajonojen, jotka voisivat tulla kyseeseen politiikalla

247

Arvojen iterointi

- * Optimaalisen politiikan selvittämiseksi lasketaan tilojen hyödyt ja käytetään niitä optimaalisen toiminnon valitsemiseen
- * Tilan hyödyksi lasketaan sitä mahdollisesti seuraavien tilajonojen odotusarvojen hyöty

- * Luonnollisesti jonot riippuvat käytetyistä politiikasta π

- * Olk. s_t tila, jossa agentti on kun π :tä on noudatettu t askelta

- * Huom. s_t on satunnaisuuttuja
- * Nyt

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

248

- * Tilan todellinen hyödyt $U(s)$ on $U^{\pi^*}(s)$

- * Palkkio $R(s)$ siis kuvaa tilassa s olevan lyhyen tähtäyksen hyödyllisyyttä, kun taas $U(s)$ on s :n pitkän tähtäyksen hyödyllisyys siitä eteenpäin laskien

- * Esimerkkinä maailmassamme maali-tilan lähellä olevilla tiloilla on korkein hyödyt-arvo, koska niistä matka on lyhin

| | | | |
|-------|-------|-------|-------|
| 0.812 | 0.868 | 0.912 | +1 |
| 0.762 | ■ | 0.660 | -1 |
| 0.705 | 0.655 | 0.611 | 0.388 |

249

- * Nyt voidaan soveltaa hyödyn odotusarvon maksimointia

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

- * Koska tilan s hyöty nyt on diskontattujen palkkioiden summan odotusarvo tästä tilasta eteenpäin, niin se voidaan laskea:

- ◇ Välttöm palkkio tilassa s , $R(s)$, +
- ◇ seuraavan tilan diskontatun hyödyn odotusarvo olettaen, että valitaan optimaalinen toiminto

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

- * Tämä on *Bellman-yhtälö*
- * Toimintaympäristössä, jossa on n tilaa, on myös n Bellman-yhtälöä

250

- * Bellman-yhtälöiden yht' aikaiseen ratkaisemiseen ei voi käyttää lineaaristen yhtälöryhmien tehokkaita ratkaisumenetelmiä, koska max ei ole lineaarinen operaatio
- * Iteratiivisessa ratkaisussa aloitamme tilojen hyödyjen mv. arvoista ja päivitämme niitä kunnes saavutetaan tasapainotila

$$U_{i+1}(s) \leftarrow R(s) + \gamma \max_a \sum_{s'} \mathcal{T}(s, a, s') U_i(s'),$$

- missä indeksi i viittaa iterointiin i hyödyt-arvoon

- * Päivitysten toistuva soveltaminen päättyy taatusti tasapainotilaan, jolloin saavutetut tilojen hyödyt-arvot ovat ratkaisu Bellman-yhtälöihin

- * Löydyt ratkaisut ovat yksikäsitteisiä ja vastaava politiikka on optimaalinen

251

Politiikan iterointi

- * Alkaen lähtöpolitiikasta π_0 toista
- * **Politiikan arviointi:** laske kaikille tiloille hyötyarvo $U_i = U^{\pi_i}$ politiikkaa π_i sovellettaessa
- * **Politiikan parantaminen:** perustuen arvoihin U_i laske uusi hyödyn odotusarvon maksimoiva politiikka π_{i+1} (vrt 250)
- * Kun jälkimmäinen askel ei enää muuta hyötyarvoja, niin algoritmi päättyy
- * Tällöin U_i on Bellman-päivityksen kiintopiste ja ratkaisu Bellman-yhtälöihin, joken vastaavaan politiikan π_i on olava optimaalinen
- * Äänellisellä tila-avaruudella on vain äärellinen määrä politiikkoja, jokainen iteratio parantaa politiikkaa, joten politiikan iterointi päättyy lopulta

252

5. Toiminnan suunnittelu

- * Tavoitteemme on valita sarja toimintoja, jotka johtavat agentin alkutilasta maaliin
- * Näinhän hakualgoritminkin toimivat
- * Hakualgoritmit eivät skaalaudu tosielämän suurin ongelmien
- * Tavoitteen kannalta irrelevantit toimintomahdollisuudet lisäävät ongelman kompleksisuutta
- * Tausatetämys auttaa karsimaan turhia hakupolkuja
- * Hakualgoritmille heuristinen funktio on määriteltävä ongelmakohtaisesti uudelleen ja huolellisesti

254

STRIPS-kieli

- * STRIPS = Stanford Research Institute Problem Solver
- * Klassinen suunnitteluympäristö = täysin havainnoitava, deterministinen, äärellinen, staattinen ja diskreetti
- * Tilojen esittämiseen käytetään propositionaalisia predikaattiteraalien konjunktioita
- * **Kentällä(Kone₁, Helsinki) ∧ ...**
- * Literaalit eivät saa sisältää sitomattomia muuttujia eikä funktioita
- * Muuttujien eksisten tiaalinen kvantifiointi on sallittua
- * Suljetun maailman oletuksen perusteella kaikkien mainisematomien literaalien oletetaan olevan epätosia

256

- * Toimintoa voidaan soveltaa, jos sen ennakkoehdot toteutuvat

$$At(P_1, JFK) \wedge At(P_2, SFO) \wedge Plane(P_1) \wedge Plane(P_2) \wedge Airport(JFK) \wedge Airport(SFO)$$

- * toteuttaa sijoituksella $\{p/P_1, from/JFK, to/SFO\}$ toiminnon $Fly(p, from, to)$ ennakkoehdot, joten $Frig(P_1, JFK, SFO)$ on sovellettavissa
- * Koska STRIPS-kielessä ei ole funktio-symboloja, on mikä tahansa toimintokkeenä muutettavissa propositionaalisiksi
- * Esimerkiksi lentokuljetusongelmassa, jossa koneita on 10 ja kenttä 5, toiminto $Fly(p, from, to)$ voitaisiin muuttaa $10 \times 5 \times 5 = 250$ propositionaalisiksi toiminnoksi

258

- * Koska kullakin kierroksella politiikka on kiinnitetty, niin politiikan arvioinnissa ei ole tarvetta maksimoida ylitoinintojen
- * Bellman-yhtälö yksinkertaistuu:

$$U_i(s) = R(s) + \gamma \sum_s' T(s', \pi_i(s), s) U_i(s')$$

- * Koska epälineaarista maksimoimisesta on päästy eroon, on tämä lineaarinen yhtälö
- * Lineaarinen yhtälöryhmä, jossa on n yhtälöä ja niissä n tuntematonta muuttujaa voidaan ratkaista ajassa $O(n^3)$ lineaarialgebran menetelmien
- * Kuutiollisen ajan sijaan voidaan tyytyä approksimoimaan politiikan laatua ajamalla vain tietty määrä yksinkert. arvojen iterointi askelia kehoillisten hyötyarvoitoiden saamiseksi

253

- * Suunnittelussa tietämyksen esitys mahdollistaa tapausriippumattoman heuristiikan käytön
- * Myös ongelmien jakamista (lähes) riippumattomiin osaongelmiin voidaan hyödyntää suunnittelussa
- * Suunnittelun ei tarvitse edetä lineaarisesti alkutilasta maaliin, vaan osaongelmien ratkaisuja voidaan lomittaa
- * Keskeinen tekijä suunnittelun onnistumisessa on tietämyksen esitys
- * Valtun kielen tulisi olla riittävän ilmaisuvoimainen ollakseen yleiskäyttöinen, mutta toisaalta riittävän rajoittanut tehokkaan suunnittelun takaamiseksi

255

- * Maalitalia on positiivisten literaalien konjunktio, jossa ei ole sitomattomia muuttujia: $Rikas \wedge Kuuluisa$
- * Tila s toteuttaa maalin g , jos se sisältää kaikki g :n literaalit: $Rikas \wedge Kuuluisa \wedge Onneton$
- * Toiminnot määritellään ennakkoehtojen ja vaikutusten avulla:
 $Action(Fly(p, from, to))$
 $Precond: At(p, from) \wedge Plane(p) \wedge Airport(from) \wedge Airport(to)$
 $Effect: \neg At(p, from) \wedge At(p, to)$
- * Tämä on erikoistapaus *toimintostekemasta*
- * Seuraukset voidaan ilmoittaa listamalla erikseen todeksi muuttuvat literaalit ja epätoiseksi muuttuvat literaalit

257