Luento 3: Johdatus peliteoriaan

Raul Hakli

Johdatus peliteoriaan

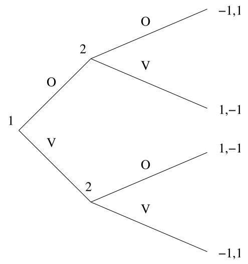
- Aiemmin käsittelimme tilanteita, joissa agentti tekee päätöksiä stokastisessa ympäristössä. Nyt siirrymme tarkastelemaan tilanteita, joissa agentteja on useita. Tällaisia ympäristöjä kutsutaan *strategisiksi* koska agentit ottavat toiset agentit huomioon päätöksenteossaan yrittämällä ennakoida niiden päätökset. Näiden tilanteiden tarkasteluun käytetään *peliteoriaa* (game theory).
- Peliteoriassa tarkastellaan itsenäisten ja omien intressiensä mukaan toimivien agenttien toimintaa ja niiden välistä vuorovaikutusta. Oletuksena on että agentit ovat rationaalisia.
- *Täydellisen informaation* peleissä agentit tietävät omat ja toistensa toimintavaihtoehdot ja utiliteettifunktiot. Lisäksi oletetaan (poikkeuksena *evolutionaarinen peliteoria* jossa näitä oletuksia ei tarvita) että kaikki tämä ja agenttien rationaalisuus on *yleistä tietoa* (common knowledge). Yleinen tieto että p tarkoittaa, että kaikki tietävät että p ja kaikki tietävät että kaikki tietävät että p jne.

Nonkooperatiivinen peliteoria

- Nyt tarkasteltavassa *nonkooperatiivisessa* (non-cooperative) peliteoriassa oletetaan, että agentit eivät voi tehdä sitovia sopimuksia (kommunikointi voidaan sallia).
- *Kooperatiivisessa* (cooperative) peliteoriassa sallitaan agenttien välinen kommunikointi ja sitovien sopimusten tekeminen.

Pelien esittäminen: ekstensiivinen muoto

Tekoälykurssilla tarkasteltiin kahden pelaajan nollasummapeleja ja esitettiin niitä pelipuina. Pelipuita kutsutaan pelin *ekstensiiviseksi muodoksi* (extensive form). Esim. Kumpi käsi?



Pelien esittäminen: normaalimuoto

Pelit voidaan esittää myös *normaalimuodossa* (normal form) eli *strategisessa muodossa*:

	O	V
O	-1,1	1,-1
V	1,-1	-1,1

- Pelien ei tarvitse olla nollasummapelejä kuten tämä vaan molemmat pelaajat voivat hyötyä samoista lopputulemista.
- Nyt ei myöskään rajoituta kahteen pelaajaan, vaan pelaajia voi olla mikä tahansa äärellinen määrä.

Normaalimuotoiset pelit

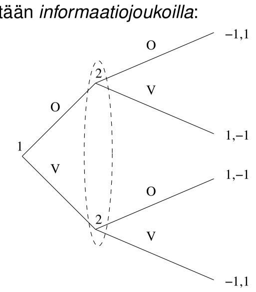
Määritelmä 1 Äärellinen, n:n agentin *normaalimuotoinen peli* on monikko (N, A, O, μ, u), missä

- $N = 1, \ldots, n$ on äärellinen joukko agentteja.
- $A = A_1, \ldots, A_n$, missä A_i on joukko agentin i *strategioita* (tai *toimintoja*). Jokaista vektoria $a = (a_1, \ldots, a_n) \in A$ kutsutaan strategia- tai toimintaprofiiliksi. Merkintä (a_{-i}) tarkoittaa kaikkien muiden agenttien toimintoja paitsi a_i:n ja merkintä (a_{-i}, a_i) tarkoittaa toimintaprofiilia jossa a_i on agentin i toiminta.
- O on *lopputulemien* joukko
- $\mu : A \to O$ on funktio toiminnoilta lopputulemille
- $u = u_1, \ldots, u_n$, missä $u_i : O \to \mathbb{R}$ on agentin i *palkkiofunktio*.

Nyt yksinkertaistetaan hieman: O=A, μ on identiteettif. ja u_i utiliteettif.

Strategioiden valinta

Normaalimuotoisissa peleissä ajatellaan pelaajien valitsevan strategiansa samanaikaisesti (toisin kuin tekoälykurssin peleissä). Tämä ei tarkoita että heidän tarvitsisi tehdä tekonsa samanaikaisesti, ainoastaan että he valitsevat tekonsa toistensa valinnoista tietämättä. Ekstensiivisessä muodossa tätä esitetään *informaatiojoukoilla*:



Rationaalisuus peleissä

- Päätöksiä ei nyt voi tehdä suoraviivaisesti utiliteetin odotusarvoa maksimoimalla: toisen agentin toimintavaihtoehdoille ei voi asettaa todennäköisyyksiä koska toinen agentti pyrkii ottamaan minun toimintani huomioon ⇒ molemminpuolinen riippuvuus.
- On kehitetty erilaisia *ratkaisukäsitteitä* (solution concepts), joiden perusteella pyritään arvioimaan, mihin lopputulemiin agentit päätyvät.
- Tarkastellaan tässä dominoitujen strategioiden poistoa ja Nash-tasapainoa.

Dominanssi

Voidaan olettaa, että rationaalinen agentti ei pelaa sellaista strategiaa, joka on joka tapauksessa (kaikilla muiden agenttien strategiaprofiileilla) aidosti huonompi vaihtoehto kuin jokin toinen strategia.

Määritelmä 2 Agentin i strategia a_i' *dominoi tiukasti* strategiaa a_i, joss

$$u_i(a_{-i}, a_i') > u_i(a_{-i}, a_i)$$

kaikilla muiden agenttien toiminnoilla a_{-i}.

Vastaavasti määritellään *heikko dominanssi* sallimalla epäyhtälössä yhtäsuuruus. Rationaalinenkin agentti voi joskus pelata heikosti dominoitua strategiaa.

Iteroitu dominoitujen strategioiden poisto

Tiukasti dominoitujen strategioiden eliminointi perustuu oletuksille, että rationaalinen agentti ei valitse tiukasti dominoitua strategiaa ja että on yleistä tietoa että kaikki agentit ovat rationaalista. Esimerkki:

	V	K	O
Y	1,0	1,2	0,1
A	0,3	0,1	2,0

Koska pelaaja 2 on rationaalinen, hän ei valitse strategiaa Oikea koska strategia Keski dominoi sitä. Pelaaja 1 tietää tämän, joten hän voi pelata ikään kuin peli olisi seuraava:

	V	K
Y	1,0	1,2
A	0,3	0,1

Iteroitu eliminointi jatkuu

Koska toisaalta pelaaja 2 tietää pelaajan 1 tietävän että pelaaja 2 on rationaalinen (ja että peli siis näyttää 1:lle tältä) ja 2 tietää että 1 on rationaalinen, hän päättelee ettei pelaaja 1 valitse strategiaa Alas, joka nyt on strategian Ylös dominoima, joten peli näyttää 2:lle seuraavalta:

	V	K
Y	1,0	1,2

Nyt 2 tietenkin rationaalisena valitsee dominoivan strategian Keski, joten tämän *iteroidun tiukasti dominoitujen strategioiden eliminoinnin* lopputuloksena saatiin tulema Y, K, jota kutsutaan *dominanttien strategioiden tasapainoksi*.

Eliminoinnista

- Pelejä, joissa prosessista jää jäljelle täsmälleen yksi strategiaprofiili, kutsutaan *dominanssiratkeavaksi* (dominance solvable) ja vastaavaa tulemaa kutsutaan pelin *ratkaisuksi* (solution). Ratkaisu on yksikäsitteinen eikä siis riipu strategioiden käsittelyjärjestyksestä.
- Kaikki pelit eivät kuitenkaan ratkea tällä tavalla. Esim.

	V	K	O
Y	1,0	1,2	0,1
A	0,3	0,1	2,2

- Ei vaadi uskomuksia toisten pelaajien käyttämistä strategioista, mutta vaatii että pelaajien rationaalisuus on yleistä tietoa.

Nash-tasapaino

Nash-tasapainolla tarkoitetaan strategiaprofiilia, jossa yhdenkään agentin ei kannata vaihtaa strategiaansa jos kukaan muukaan ei vaihda.

Määritelmä 3 Strategiaprofiili a^* on Nash-tasapaino, joss kaikille pelaajille i pätee

$$u_i(a_{-i}^*, a_i^*) \geq u_i(a_{-i}^*, a_i)$$

millä tahansa strategialla $a_i \in A_i$.

Toisin sanoen Nash-tasapaino on toimintaprofiili, jossa kaikkien pelaajien strategiat ovat *parhaita vastauksia* toisten pelaajien strategioihin. Tällaisen tilan sanotaan olevan vakaa l. stabiili, koska kenelläkään pelaajista ei ole motivaatiota poiketa omasta strategiastaan.

Edellisen kalvon pelissä on yksi Nash-tasapaino: (Y,K).

Nash-tasapaino ja dominanssi

- Iteroitu (tiukasti) dominoitujen strategioiden eliminointi ei koskaan poista pelistä sellaista tilaa joka on Nash-tasapaino.
- Jos poistettaisiin heikosti dominoituja strategioita, niin myös Nash-tasapainotiloja voisi kadota.
- Jos peli on dominanssiratkeava, niin jäljelle jäänyt tila on pelin ainoa Nash-tasapaino.

Sekastrategiat

Nash-tasapainon ongelmia:

1. joissain peleissä niitä on useita → vaikea ennustaa mihin päädytään (tai mikä pitäisi valita)
2. joissain peleissä niitä ei ole lainkaan kun rajoitutaan puhtaisiin strategioihin → käytettävä sekastrategioita

Määritelmä 4 *Sekastrategia* (mixed strategy) pelaajalle i, jonka puhtaat strategiat ovat $A_i = \{a_{i1}, \ldots, a_{ik}\}$, on todennäköisyysjakauma $\sigma_i = \{p_{i1}, \ldots, p_{ik}\}$, missä $\forall j = 1, \ldots, k \; 0 \leq p_{ij} \leq 1$ ja $\sum_{j=1}^{k} p_{ij} = 1$.

Sekastrategiat voidaan tulkita joko satunnaistetuksi valinnaksi puhtaiden strategioiden välillä tai epävarmuudeksi toisen pelaajan käyttäytymisestä.

Sekastrategiat (2)

Esim. seuraavassa pelissä ei ole Nash-tasapainoa puhtailla strategioilla.

	O	V
O	-1,1	1,-1
V	1,-1	-1,1

Kun sallitaan sekastrategiat, löydetään Nash-tasapaino $((\frac{1}{2}, \frac{1}{2}), (\frac{1}{2}, \frac{1}{2}))$.

Lause 1 (Nash 1950) *Jokaisella äärellisellä strategisella pelillä on Nash-tasapaino.*

- Kuinka tasapaino löydetään?
- Kuinka valitaan useista tasapainoista?

Pareto-optimaalisuus

Pelin lopputulemia voidaan myös arvioida niiden optimaalisuuden perusteella. Optimaalisuudella tarkoitetaan tässä Pareto-optimaalisuutta, joka määritellään seuraavasti:

Määritelmä 5 Toimintaprofiili a on *Pareto-optimaalinen*, joss ei ole olemassa toista toimintaprofiilia a' s.e. jollekin agentille i $u_i(a') > u_i(a)$ ja kaikille muille agenteille $u_j(a') \geq u_j(a)$.

Pareto-optimaalinen tila on siis sellainen, josta kenenkään utiliteettia ei voi parantaa huonontamatta samalla jonkun muun utiliteettia. Pareto-optimaalisuus ei ole varsinainen ratkaisukäsite (koska pelaajat eivät tavoittele Pareto-optimaalisuutta vaan maksimoivat omaa utiliteettiaan) mutta sen avulla voi vertailla eri lopputulosten hyvyyttä.

Vangin dilemma

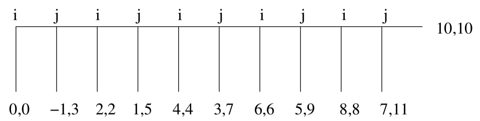
Esimerkki: Vangin dilemma (Prisoner's Dilemma, PD):

	C	D
C	3,3	1,4
D	4,1	2,2

- Jos peliä pelataan kerran, kumpi pitäisi valita: C vai D?
- Peliteorian mukaan rationaaliset agentit valitsevat D:n.
- Tässä agentin 1 preferenssit ovat seuraavat: $D, C \succ_1 C, C \succ_1 D, D \succ_1 C, D$. D dominoi tiukasti C:tä ja D, D on Nash-tasapaino.
- Kuitenkin tulema D, D on alioptimaalinen, koska C, C olisi kummallekin aidosti parempi vaihtoehto!

Iteroitu PD ja Centipede

- Jos agentit pelaavat vangin dilemmaa toistuvasti, agentit voivat tehdä päätöksiä sen mukaan mitä on tapahtunut aiemmilla kierroksilla → syntyy mahdollisuus rankaista agentteja jotka eivät tee yhteistyötä.
- Yhteistyön tekeminen muuttuu houkuttelevammaksi, koska pitkällä tähtäimellä hyötyy enemmän ja vaikka alussa joutuisikin "petetyksi", tappio ei ole pitkällä aikavälillä suuri.
- Tosin jos peliä pelataan vain äärellisen monta kertaa ja lukumäärä on agenttien tiedossa, viimeisellä kierroksella pettäminen dominoi.
- Vrt. tuhatjalkainen (centipede) on peli, jossa rationaalinen agentti lopettaa heti ensimmäisen tilaisuuden tullen.



Axelrodin turnaus

Axelrod järjesti vuonna 1980 turnauksen, jonne kutsui eri tieteenalojen edustajilta tietokoneohjelmia, jotka pelaisivat iteroitua vangin dilemmaa.

ALL-D Valitse aina D.

RANDOM Valitse satunnaisesti.

TIT-FOR-TAT Aloita C:llä ja sen jälkeen matki aina toista.

TESTER Aloita D:llä. Niin kauan kuin toinen valitsee C:n, valitse aina kaksi kertaa C ja sitten D. Jos toinen joskus valitsee D:n, niin jatka toistamalla toisen edellistä tekoa kuten TIT-FOR-TAT,

JOSS Toimi muuten kuin TIT-FOR-TAT, mutta muuta keskimäärin yksi kymmenestä C:stä D:ksi.

Kukin ohjelma pelasi jokaista muuta vastaan viisi peliä, joista kukin koostui kahdesta sadasta kierroksesta. Voittajaksi valittiin yhteensä eniten utiliteettipisteitä kerännyt strategia. Parhaiten menestyi TIT-FOR-TAT.

Lähteitä

- Martin J. Osborne & Ariel Rubinstein: *A Course in Game Theory*, MIT Press, 1994.
- Vlassis, luku 3.
- Wooldridge, luku 6.