

# TEKOÄLY

## syksy 2004

4 ov, jatko-opinotekelinen (?)

Luennot ti 12–14 S2, to 12–14 TB109  
31.8 – 9.12. (14 viikkoa)

syysloma 18.–24. 10.

prof. Tapio Elomaa

Harjoitustyöt 4 kpl

Viikkoharjoitukset

ke 10–12 CI03, DI Teppo Soiminen  
to 9–11 B206, arkkit. yo. M. Ruuska

Koe 14. joulukuuta (alustava tieto)

Koe 30 p. + ht: 24 p. + laskarit 6 p. =  
60 p.

{ elomaa, ruuska2, teppos }@cs.tut.fi

1

## Viikkoharjoitukset

- \* Viikkoharjoituksiin osallistuminen on erittäin suositeltavaa
- \* Osallistumista saa "merkinnän". Valmiudesta esittää taululla vastaus kysymyksen saa kustakin merkinnän
- \* Kussakin harjoituksissa on n. 5 tehtävää ⇒ kaikkiaan merkintöjä voi kerätä n. 6 · 12 = 72 kappaletta

Osuus	Lisäpisteet
merkinnöistä	kurssin arvostelussa
25%	1
40%	2
55%	3
70%	4
80%	5
90%	6

3

## Luentoaikataulu

1. **Johdanto** (I, 1)
  - \* Taustaa
  - \* Tekoälyn historia
  - \* Älykkäät agentit
2. **Logiikka, tietämys ja päättely** (III, 2–3)
3. **Ongelmanratkaisu ja hakualgoritmit** (II, 4–5)
4. **Epävarmu tietämys ja päättely** (V, 6–8)
5. **Toiminnan suunnittelu** (IV, 9)
6. **Koneoppiminen** (VI, 10–12)
  - \* Päätöspuiden oppiminen
  - \* Hypoteesikokeelmien oppiminen
  - \* Laskennallinen oppimisteoria
  - \* Palautteoppiminen
  - \* Tilastollinen oppiminen
7. **Muut mahdolliset asiat** (13–14)

5

7

## Harjoitustyöt

Kurssilla on neljä pakollista harjoitustyötä. Kurssin suorittaminen edellyttää kaikkien töiden tekemistä

"Hankugeneraattori"

(+selostus 1–2 sivua). Max 5 p.

*Palautuspäivä* 1. 11.

Trimok-agentin käyttämien tekoälytekniikoiden selvitys (n. 5 sivua). Max 6 p.  
*Palautuspäivä* 1. 12.

Trimok-peliä pelaava agentti. Max 6 p.  
Turnauksen 3:lle parhaalle lisäpisteet 3–2–1.  
*Palautuspäivä* 15. 12.

Essee. Vapaaehtoinen aihe oppiäksen aihealueelta (n. 10 sivua). Max 7 p.  
*Palautuspäivä* 15. 12.

2

## Materiaali

Kurssin oppikirja on:

- \* S. Russell, P. Norvig: *Artificial Intelligence, A Modern Approach*, Second ed., Prentice Hall, 2003

Valmistaa monistetta ei ole, luentokalvot pannaan verkkoon luentojen tahiin.

www.cs.tut.fi/kurssit/8101905/

www.cs.tut.fi/~elomaa/opetus/AI04/

4

## Määritelmät

- \* Älykkäälle koneelle ainoa tuntemamme vertailukohta olemme me itse
- \* Toisaalta vertailu *inhimilliseen* älykkyyteen rajaa pois muut (paremmat) vaihtoehdot
- \* Ideaalista älykkyyttä kutsutaan *rationaalisuudeksi*
- \* Toisaalta älykkyyttä voidaan tarkastella *ajattelun* tai *toiminnan* näkökulmasta
- \* Yhdistelmänä saadaan neljä toisistaan poikkeavaa näkökulmaa tekoälyyn

7

## 1. Johdanto

- \* Tekoäly on käsitteenä laaja, korkealentoinen ja ajan myötä muuttuva
- \* Nykyisin ehkä enemmän filosofien ja kognitotieteilijöiden kenttää
- \* Tietojenkäsitteellisesti tekoäly käsittää joukon tavoitteelliaan rajatunpia tutkimusaloja, jotka ovat jo eriytyneet melko etälle toisistaan
- \* Yhteistä eri osa-alueille on pyrkimys koneiden "älykkyyden" nostamiseen
  - ◊ (Toisin katsoen: ohjelmistojen käytön helpottamiseen)
- \* Tuloksena käytökeelpoisia ohjelmistoja ja mekaanisen laskennan rajoja karttettavaa teoriaa

6

## Turingin testi

- \* Englantilaisen matemaatikon Alan Turingin 1950 esittämä kriteeri koneen älykkyydelle on se, ettei ihmisen erota konetta ihmisestä keskustellessaan sen kanssa tekstiviestein
- \* Eräs inhimillisen toiminnan testi
- \* Ns. *totalisessa* Turingin testissä koneen on pystyttävä myös havainnoimaan ja manipuloimaan fyysisistä ympäristöään
- \* Atkarajoitteita kilpailuja järjestetään vuosittain
- \* Parhaiten knoppitetoisia kilpailuttajia vastaan pärjävät huijausohjelmat
- \* Inhimilliset asiantuntijat arvioidaan suurimmalla trillä koneiksi

8

## gnitiivinen mallintaminen

Inhimillisen ajattelun tutkimus kuuluu psykologian ja kognitiotieteeseen

Myös tietysti neurofysiologia jne.

Vahva yhteys tekoälytutkimukseen, jos näkökulma on inhimillinen ajattelu

Ajattelun/mielen algoritmikka ja tietämysrakenteet

Allen Newell ja Herbert Simon (1961): General Problem Solver (GPS) –päättelyskelman ja inhimillisen päättelyn vertailu

9

\* Rationaalinen agentti toimii parhaan mahdollisen lopputuloksen saavuttamiseksi havaintojensa ja tietämyksensä puitteissa

\* Epävarmuuden vallitessa pyritään maksimoimaan odotusarvoa

\* Rationaalinen toiminta pitkällä tähtäyksellä voi edellyttää epärationaalista vaikuttavaa toimintaa lyhyellä tähtäyksellä

\* Reaalimaailmassa täydellinen rationaalisuus ei useimmiten ole mahdollista (ajampuuteen takia)

## Tekoailyn historiaa

\* Ensimmäiseksi tekoälyjulkaisuksi voidaan katsoa McCulloch ja Pitts (1943)

\* Esitetään kuinka yksinkertaisten laskentaelementtien, *neuronien*, muodostamalla verkolla voidaan laskea loogiset konnektiivit (**and**, **or**, **not**, jne.)

◊ Osoitetaan, että kaikki laskettavat funktiot voidaan laskea neuroverkolla

◊ Vinjattiin verkkojen voivan oppia

\* Hebb (1949) antaa yksinkertaisen päivityssäännön neuroverkkojen opettamiseen

13

\* Robotiikkakin eteni: esim. SRI:n Shakey (1969)

\* Samoin neuroverkko tutkimus (Widrow & Hoff, Rosenblatnin perceptron)

\* Vähitellen kuitenkin kävi selväksi, että mikromaailmojen menestys ei sellaisenaan skaalautunut ylöspäin. Se oli saavutettu ilman syvempää ymmärtämystä koneongelmaista ja laskentaintensiivisin menetelmin

\* Neuroverkko tutkimuksen tietojenkäsittelystä tappoi yli vuosikymmeneksi Minskyn ja Papertin todistus perceptronin heikosta esitysvoimasta (xor-funktio)

\* 1970-luvulla kehitettiin asiantuntijajärjestelmiä, joihin kerättiin yhden sovellusalueen syvää tietämystä

15

## tionaalisuus

Rationaalinen ajattelu on oleellisesti formaalin logiikan ja loogisen päättelyn tutkimista

(Yksinomaan) logiikkaan perustuvat menetelmät käsitvät laskennan vaativuudesta sekä epävarman tiedon esittämisen vaikeudesta

Rationaalisen toiminnan mallissa tarkastellaan *agentteja*

Agentti on toimija

Ohjelmistoagenttiin erottaa ohjelmasta esim. sen autonomisuus, ympäristön havainnointikyky, adaptiivisuus, pitkä elinikä, kyky omaksua muiden tavoitteita, jms.

10

## Taustatietoita

\* Filosofia ja matematiikka

◊ Formaali päättely, mielen ja tiedon olemukset, jne.

◊ Päättelysäännöt, laskettavuus ja sen vaativuus, epävarma päättely (todennäköisyyslaskenta), jne.

\* Taloustiede

◊ päätös- ja peliteoria, operaatio tutkimus, jne.

\* Neurotieteet

\* Psykologia ja kognitiotiede

\* Tietotekniikka

\* Kontrolliteoria ja kybernetiikka

\* Lingvistiikka

12

\* Turing (1950) esittelee testinsä, koneoppimisen, geneettiset algoritmit ja palauteoppimisen

\* 1956 John McCarthy järjesti aihealueesta kiinnostuneiden tutkijoiden tapaamisen, missä tutkimusalue sai nykyisen nimensä

\* Heti alusta lähtien keskeisiä ylipistoja olivat nykyisinkin tekoälytutkimuksen huipulla olevat CMU, MIT ja Stanford

\* McCarthy (1958) Lisp-ohjelmointikieli

\* 1950- ja 1960-luvulla saavutettiin huomattavia edistysaskelia mikro-maailmoissa (esim. palikkamaailma) tapahuuvasa toiminnassa

14

\* Asiantuntijajärjestelmillä saavutettiin monilla aloilla ihmisasiantuntijoita parempi suorituskyky ja niistä tuli tekoälytutkimuksen ensimmäinen kaupallinen menestystarina

\* Asiantuntijajärjestelmien kehittämisen osoittautui kuitenkin vaativaksi käsitöksi

\* Logiikkaohjelmointi eli kirkkaimman kukoistuksensa 1980-luvun puollessa välissä

\* Neuroverkko tutkimus nousi taas tutkimusalaaksi tietojenkäsittelyssä 1980-luvun puolivälisistä lähtien

\* Myös koneoppimistutkimuksen nousukausi ajoittuu 1980-luvulle

16

Samaten Bayes-verkkojen tutkimus lähti liikkeelle tuolloin

- ◊ Ehkä toinen merkittävä kaupallisen sovellus Microsoftin vahvan näkemyyden takia

Sittenmin samoja asioita on tutkittu myös tietämyksen muodostamisen nimikkeellä

Agenttitekniologia on nouseva trendi monella tietojenkäsittelyn alueella

- Ylimeaikainen kehitys on ollut myös suuntautuminen analyttiseen tutkimukseen *ad hoc* tekniikkojen sijaan
- ◊ Koneoppimisen teoreettiset mallit
- ◊ Toiminnan suunnittelun uudet menetelmät
- ◊ Peliteorian uusi kukoistus

17

## 1.1 Älykkäät agentit

Agentti havainnoi *toimintaympäristään sensorein* ja vaikuttaa siihen *aktuatoorin*

- ◊ Ihmisen sensoreita ovat mm. silmät, korvat ja nenä sekä aktuatooreita esim. kädet ja jalat
- ◊ Robotti sensoreita voivat puolestaan olla kamerat ja laseretäisyysmittarit, aktuatooreita ovat eri moottorit
- ◊ Ohjelmistoagentti puolestaan havainnoi näppäinpainalluksia, tiedostojen sisältöjä ja tietoliikennepaketteja
- ◊ Sen toimintoja puolestaan ovat näyttölle tai tiedostoon kirjoittaminen ja tietoliikennepaketien lähettäminen

18

minen ja tietoliikennepaketien lähettäminen

- \* Oletamme agentin havainnoivan oman toimintansa, muttei välttämättä niiden vaikutusta

- \* Yleisesti ottaen agentin toiminta kullakin heikellä voi riippua kokosen havaintohistoriasta

- \* *Agenttifunktio* kuvaa havaintojonon toiminnaksi

- \* Funktion kaikkien mahdollisten syöte-vaste-parien taulukko on täydellinen ulkoinen kuvaus agentista

- \* Sisäisesti agentin määrittää sen kontrolliohjelma

19

### Agentin toiminnan arviointi

- \* Tavoitelemamme rationaalisen agentin pitäisi menestyä suorittamassaan tehtävässä

- \* Menestyksen arvioimiseksi tarvitsemme *tuloksellisuustietoa*

- \* Toiminnan rationaalisuus riippuu
  - ◊ tuloksellisuustiestä joka määrää menestyksen,

- ◊ agentin maailmaa koskevista ennakkotiedoista,

- ◊ sen mahdollisista toimintoista ja
- ◊ havaintohistoriasta

Jokaisella havaintojonolla *rationaalinen agentti* valitsee toiminnan, joka odotusarvoisesti maksimoi tuloksellisuustietan havaintohistorian ja ennakkotietojen valossa

20

## Toimintaympäristöjen ominaisuuksia

Toimintaympäristöjä voidaan luokitella ainakin seuraavien ominaisuuksien perusteella

- \* **Täysin vai osittain havainnoitava?**

- ◊ Täysin havainnoitavassa maailmassa agentti saa sensoreiltaan kaiken toiminnan valintaan vaikuttavan relevantin tiedon

- ◊ Näin ollen se ei tarvitse ylläpitää omaa käsitystään maailman tilasta
- ◊ Ympäristö voi olla vain osittain havainnoitavissa perustuloiteensa takia tai sensoreiden epä tarkkuuden vuoksi

21

- \* **Deterministinen vai stokastinen?**

- ◊ Jos maailman nykyinen tila ja agentin suorittama toiminto määräävät seuraavan tilan, on maailma deterministinen. Muuten se on stokastinen.

- ◊ Deterministinenkin maailma voi vaikuttaa stokastisella, jos se ei ole täysin havainnoitavissa

- \* **Episodinen vai sekventiaalinen?**

- ◊ Episodisessa maailmassa edellisten episodien toiminnot eivät vaikuta tulevisissa episoodeissa
- ◊ Sekventiaalisessa maailmassa puolestaan tehtävät päätökset vaikuttavat myös tuleviin päätöksiin

22

- \* **Staatinen vai dynaaminen?**

- ◊ Staatisessa maailmassa agentti voi pysähtyä pohittamaan toimintansa ilman pelkoa asetelman muuttumista

- ◊ Dynaamisessa maailmassa agentin on jatkuvasti tarkkailtava maailman tilaa

- ◊ *Semidynaamisessa* ympäristössä maailma sinänsä ei muutu ajan kuluessa, mutta agenttia rangaitaan toiminnan suunnitteluun kuluvasista ajasta

- \* **Diskreetti vai jatkuva-arvoinen?**

- ◊ Jako kohdistuu maailman tilaan, aikaan ja agentin havaintoihin ja toimintoihin

23

- \* **Yksi vai monta agenttia?**

- ◊ Moneen agentin maailmassa voidaan kilpailia tai tehdä yhteistyötä

- \* Esimerkiksi robotin kannalta "reaalimaailma" on

- ◊ *Osittain havainnoitava*  
Sensorit ovat epätäydellisiä ja havainnoivat vain lähiympäristöä (sama pätee myös ihmiselle)

- ◊ *Stokastinen*

- Pyörät lipsuvat, akut tyhjenevät, osat hajoavat — koskaan ei ole varmaa, että aiottu toiminto toteutuu

24

#### ◊ *Sekkentiaalinen*

Toimintojen seuraukset muuttuvat ajan myötä ⇒ robotin on hallittava sekventiaalisia päätösongelmia ja kyettävä oppimaan

#### ◊ *Dynaaminen*

Milloin pohtia, milloin toimia

#### ◊ *Jalkura/Ääretön*

Algoritmien on toimittava tässä ympäristössä, ei esimerkiksi äärellisessä diskreetissä hakuavaruudessa

#### ◊ *Yhden/monen agentin maailma*

Rippuu siitä halutaanko muut olot nähdä agenteina vai sitokasitesti toimivina ympäristön osina

25

### ◊ *Refleksiivinen agentti*

\* Yksinkertaisin kontrolliohjelma määrää agentin toimimaan tämänhetkisen havainnon perusteella jättäen aiemmat havainnot huomiotta

\* Refleksit ovat käytössä hätäkeinoina niin ihmisillä kuin robotillaikin

\* Refleksiivinen toiminta tuottaa oikeita päätöksiä vain jos toimintaympäristö on täysin havainnoitavissa

\* Agentin sisäinen tilan perusteella voidaan valita kulloinkin voimassa oleva sääntöjoukko

\* Täten pystytään ylläpitämään maailmanmallia, joka kattaa osia siitä, mitä ei voida havainnoida

27

### ◊ *Hyötyjä tavoitteleva agentti*

\* Agentti voi saavuttaa tavoitteensa monella tapaa, ratkaisulla voi olla laatueroja

\* Tavoitteen asettaminen sellaisenaan ei riitä ilmaisemaan monimutkaisia aseteimia

\* Jos maailman mahdollisille tiloille asetetaan järjestys *hyötyfunktioilla* (utility  $f$ ), niin agentti voi pyrkiä parantamaan sen arvoa

\* Hyötyfio kuvaa tilan (tai tilajonon) reaaliarvoksi

\* Funktiota ei tarvitse olla eksplisittisesti olemassa, jotta menetelmää voitaisiin käyttää

29

### ◊ *Tavoitehakuinen agentti*

\* Agentilla on havaintojensa lisäksi tiedossaan tavoite, johon se pyrkii

\* Tavoite on jokin ympäristötä koskeva väittäjä, joka tulisi toteuttaa

\* Yhdistämällä tavoite ja tieto mahdollisten toimintojensa vaikutuksista voi agentti pyrkiä toteuttamaan tavoitteen

\* Jos tavoitetta ei voida saavuttaa suoraan yhden toiminnon seurauksena, niin on suunniteltava sarja toimintoja sen saavuttamiseksi

\* Voidaan käyttää joko hakualgoritmeja (luentovaikeut 5–6) tai toiminnan suunnittelua (7)

26

### ◊ *Oppivat agentit*

\* Agenttien koodaaminen käsin kaikkiin mahdollisiin tehtäviin vaikuttaa toivottamalta tehtävältä

\* Jo Turing (1950) ehdotti oppimista menetelmäksi älykkäiden järjestelmien luomiseksi

\* Agentiteknologiassa oppimiskomponentin on oltava irrallinen varsinaisesta toimintakomponentista

\* Palaamme oppimisen tekniikoihin yksityiskohtaisesti kurssin loppupuolella

30