Luento 1: Johdatus hajautettuun tekoälyyn

Raul Hakli

Kurssin sisältö

- Kurssi antaa perustiedot hajautetun tekoälyn (Distributed Artificial Intelligence, DAI) keskeisistä tavoitteista, ongelmista ja menetelmistä.
- Kurssi jatkaa Tekoäly-kurssin aihepiiriä tarkastelemalla älykkään agentin käsitettä ja toimintaa sellaisissa tilanteissa, joissa agentteja on lähtökohtaisesti useita.
- Käsiteltäviä aihepiirejä ovat esimerkiksi agenttien intentionaalisten tilojen esittäminen, modaalilogiikka, peliteoria ja mekanismien suunnittelu sekä monen agentin palauteoppiminen.
- Sellaisia asioita ei käsitellä, joita käsitellään laitoksen muilla säännöllisesti järjestettävillä kursseilla kuten Tekoäly, Koneoppiminen, Hajautetut järjestelmät, Ohjelmistoagenttiteknologia.

Kurssin aikataulu

Luennot: 10.03.–11.04. MA 11–14 D122, TO 9–12 C222 Yonleht. Raul Hakli.

Harjoitusryhmät: 14.03.–22.04. TI 10–12 DK118 FM Olli Niinivaara.

Kurssikoe: TI 26.4. klo 14–17 CK112

Tapaaminen: KE 11.5. klo 13–16 C222 ???

Kurssimateriaali

Kurssilla ei käytetä varsinaista oppikirjaa, mutta perustiedot löytyvät tiiviissä muodossa Nikos Vlassisin monisteesta "A Concise Introduction to Multiagent Systems and Distributed AI". Muita hyödyllisiä kirjoja ovat seuraavat:

- tekoälyn perusoppikirja: Stuart Russell ja Peter Norvig: Artificial Intelligence: A Modern Approach (Second Edition). Prentice Hall, 2003,
- hajautetun tekoälyn oppikirja: Michael Wooldridge: An Introduction to MultiAgent Systems, John Wiley & Sons, 2002 ja
- kokoomateos hajautetusta tekoälystä: Gerhard Weiss (Ed.) Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence, MIT Press, 1999.

Arvostelu

- Kurssin suoritus vaatii kurssikokeen suorittamista ja esseen tai harjoitustyön tekemistä. Luennot ja laskuharjoitukset ovat vapaaehtoisia. Kokeessa tentitään luennot ja sovittu lisämateriaali.
- Tekemällä kurssiin liittyviä töitä ansaitsee kurssipisteitä.
- Pisteitä saa luennoista (1/läsnäolo + 1/oppimispäiväkirja, max n. 16), laskuharjoituksista (1–2/tehtävä, max n. 20), artikkelireferaateista (2–3/journal-artikkeli, 1/konferenssiartikkeli, max n. 10–12), kokeesta (30) ja esseestä/harjoitustyöstä (40).
- Arvosanarajoja tai maksimipisteitä ei ole määritelty etukäteen. Parhaan arvosanan raja asettunee johonkin välillä 90–100. Suoritukseen vaadittaneen n. 50, (tavallisesti noin puolet korkeimmasta ansaitusta pistemäärästä).
- Oppimispäiväkirjat palautetaan saman päivän aikana. Artikkelireferaatteja voi tehdä viimeisiin laskuharjoituksiin saakka.

Mitä on hajautettu tekoäly? (1)

Kaksi tapaa ajatella:

1. ”Älykkyyden” hajauttaminen: älykästä toimintaa syntyy yksittäisten laskentayksiköiden vuorovaikutuksesta. Yksiköt itse eivät ole älykkäitä vaan voivat toimia hyvin yksinkertaisten toimintasääntöjen perusteella.
2. Laskentayksiköt ovat itsekin älykkäitä $\rightarrow$ autonomiset agentit jotka tekevät itsenäisiä päätöksiä.

Varsinkin DAI:n alkuvaiheessa 1. oli suosittu lähestymistapa. Hajautus oli tapa ratkaista monimutkaisia ongelmia hajottamalla ne osaongelmiin ja kokoamalla sitten tulokset yhteen. Sittemmin agentit ovat tulleet monimutkaisemmiksi ja ”älykkäiksi” $\rightarrow$ 2.

Mitä on hajautettu tekoäly? (2)

- Kuitenkin 1. on joillakin alueilla vallitseva suunta, esim. keinoelämä (artificial life), parviäly (swarm intelligence), sosiaalinen simulaatio (social simulation).
- Myös hajautetussa tekoälyssä jotkut ovat sitä mieltä että agenttien tulisi olla mahdollisimman yksinkertaisia ja älykkään toiminnan tulisi olla *emergenttiä* eikä valmiiksi ohjelmoitua. (Mitä se oikeastaan tarkoittaa?)

Kaksi suuntausta (1)

Toinen tapa jaotella lähestymistapoja on sen mukaan, mistä agentit saavat tavoitteensa.

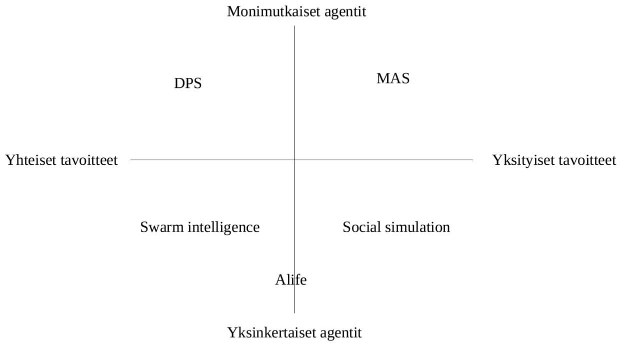
1. Hajautettu tekoäly keskittyi aluksi *hajautettuun ongelmanratkaisuun* (distributed problem solving, DPS, cooperative problem solving, CPS, CDPS). Siinä järjestelmä toimii useista laskentayksiköistä, jotka kaikki kuitenkin toimivat saman ongelman ratkaisemiseksi. Ei keskinäistä kilpailua vaan algoritmiset ongelmat koskevat lähinnä tehtävien ja resurssien mahdollisimman tehokasta jakoa.

Kaksi suuntausta (2)

2. Sittemmin on keskitytty *moniagenttijärjestelmiin* (multi-agent systems, MAS), joissa laskentayksiköt ovat autonomisia agentteja, jotka toimivat toimivat omien tavoitteidensa perusteella (self-interested). Ei siis tehdä oletusta siitä että agentit ovat hyväntahtoisia ja pyrkivät toteuttamaan jonkin yhteisen päämäärän, vaan eri agenttien yksityiset tavoitteet voivat olla keskenään ristiriitaisia, jolloin resursseista syntyy kilpailua agenttien välillä.

Aiemmin CDPS ja MAS ajateltiin erillisiksi, nykyään ehkä CDPS ajatellaan MAS:n osa-alueeksi.

Jaottelua



Kurssin aihe

Tekoälykurssilla käsittelimme rationaalisia agentteja jotka pyrkivät toimimaan optimaalisesti toteuttaakseen tavoitteitaan. Palautetaan mieliin myös muut vaihtoehtoiset lähestymistavat.

Nelikenttä:

	Rationaalinen	Ihmismäinen
Ajattelu		
Toiminta		

Lähestymistapa ja tavoitteet riippuvat toisistaan.

Rationaaliset agentit

- Tällä kurssilla jatkamme rationaalisten agenttien käsittelyä, mutta emme enää rajoitu tilanteisiin, joissa on vain yksi agentti. Agentteja voi olla useita ja se mitä järjestelmässä tapahtuu, riippuu kaikkien toiminnasta.
- Lisäksi hajautetun tekoälyn järjestelmät ovat luonteeltaan sellaisia että järjestelmään vaikuttavat agentit voivat olla joko ohjelmistoagentteja tai ihmisiä eikä agentin luonne välttämättä ole suoraan tiedossa.
- Väitän että tästä seuraa ettemme voi rajoittua vain rationaalisten agenttien toiminnan tarkasteluun, vaan joudumme myös (ainakin jossain määrin) tarkastelemaan ihmismäistä päätöksentekoa.

Agentin käsite

Heikossa mielessä agentiksi voidaan ymmärtää mikä tahansa olio, joka havaitsee ympäristöään ja vaikuttaa siihen teoillaan (Russell & Norvig, 2003).
Vahvassa mielessä agentilta vaaditaan usein seuraavia:

- sijoittuneisuus (situatedness) ympäristöön: ympäristö vaikuttaa agenttiin ja agentti ympäristöön
- autonomia: agentti päättää itse tekemisistään (vrt. olio)
- reaktiivisuus: agentti reagoi ympäristön muutoksiin
- proaktiivisuus: agentti tavoittelee päämääriään järkevästi (sitoutumisen taso), tekee aloitteita ja ennakoi käyttäjän tarpeita
- oppiminen?
- sosiaalisuus?

Sukulaistieteenaloja

- tietojenkäsittelytiede, erit. hajautetut järjestelmät
- psykologia ja kognitiotiede
- yhteiskuntatieteet, erit. taloustiede ja sosiologia
- filosofia, erit. toiminnan filosofia
- matematiikka ja logiikka

Eroja tavanomaisiin järjestelmiin (1)

- Järjestelmä koostuu useista *agenteista*. Voivat olla homo- tai heterogeenisiä, robotteja tai ohjelmistoagentteja
- Agentit ovat sijoitettuna *ympäristöön*, joka voi olla ohjelmallinen tai fyysinen. Ympäristö ei yleensä ole *staattinen* vaan *dynaaminen* jo siksi että siellä on useita agentteja.
- Agentit *havaitsevat* ympäristöään, joka voi olla kokonaan tai osittain havaittava.

Eroja tavanomaisiin järjestelmiin (2)

- Järjestelmän *kontrolli* on hajautettu: ei yhtä prosessia (tai prosessoria) joka keräisi yhteen agenttien keräämän informaation ja sitten kertoisi kullekin mitä pitää tehdä seuraavaksi.
- Järjestelmän *tietämys* on hajautettu. Ei yhteistä tietovarastoa, vaan agentit keräävät informaatiota ja päättelevät oman ja toisten agenttien tiedon perusteella.
- Agenttien on *kommunikoitava* keskenään voidakseen vaihtaa tietoa ja tehdä yhteistyötä. Vaatii kommunikaatiokielen ja -protokollan.

Etuja

Usein mainittuja moniagenttijärjestelmien etuja:

- *Rinnakkaisuus* voi tehostaa järjestelmän toimintaa (toisaalta kommunikointi, monimutkaisuus).
- Järjestelmän *vakaus* ja *luotettavuus* on usein parempi, koska jonkin agentin epäonnistuminen ei (välttämättä) kaada koko järjestelmää.
- *Skaalautuvuus* ja *joustavuus*: uusia agentteja on (pitäisi olla) helppo lisätä järjestelmään.
- *Kustannukset* voivat olla pienempiä (?)
- *Kehittäminen* ja *uudelleenkäyttö* helpompaa modulaarisuuden ansiosta.

Motivaatiota

- Tekoäly: Nykyinen suuntaus tekoälyssä: älykäs agentti. Pyritään jonkinlaiseen synteesiin, jossa aiemmin erilliset alueet yritetään yhdistää älykkääksi agentiksi.
- Tekoäly ja kognitiotiede: Älykkyys on sosiaalista. Ihmiset oppivat vuorovaikutuksessa muiden ihmisten kanssa.
- Tietojenkäsittely yleensä: "There is no such thing as a single-agent system".
- Sovellukset: Internet, sähköinen kaupankäynti, liiketoimintaprosessit, hajautetut sensorit, tiedonhaku, vertaisverkot, käyttöliittymät ja personointi, pelit ja virtuaaliympäristöt, sosiaalinen simulointi jne

Haasteita ja ongelmia (1)

- Alkuperäisen ongelman jakaminen osatehtäviksi
- Hajautetun kontrollin toimivuus
- Koordinaatiomekanismin toimivuus
- Monen agentin suunnittelu- ja oppimisalgoritmit
- Hajautetun tietämyksen hallinta ja hyväksikäyttö toisaalta järjestelmän tasolla, toisaalta yksittäisten agenttien tasolla
- Kommunikaatiokielen ja -protokollan luominen, agenttien välisen kommunikaation hallinta

Haasteita ja ongelmia (2)

- Agenttien välinen neuvottelu ja konfliktien selvittäminen. Valehtelun ja hyväksikäyttämisen estäminen.
- Agenttien sosiaalinen toiminta, normit ja konventiot, yhteistyö, ryhmiät, organisaatiorakenteet, roolit.
- Kuinka järjestelmä kokonaisuutena saadaan toimimaan koherentisti ja vakaasti?
- jne

Luentojen sisältö (1)

Suunnitelma:

To 10.3. Johdanto: kurssin hallinnolliset asiat ja johdatus aihepiiriin.

Ma 14.3. Älykkäät agentit, agenttiarkkitehtuurit: Mikä on agentti, mitä on sen sisällä ja kuinka se toimii?

Ti 15.3. Harjoitukset 1 (ex tempore)

To 17.3. Agenttien vuorovaikutus: Mitä jos agentteja on useita? Peliteorian alkeet.

Ma 21.3. Agenttien yhteistyö: Koordinaatio, kooperaatio, konventiot, neuvottelu, argumentointi.

Ti 22.3. Harjoitukset 2 (luennoista 14.3 ja 17.3.)

24.3.-30.3. Pääsiäisloma

Luentojen sisältö (2)

To 31.3. Tietämys: Agenttien mentaalisten tilojen esittäminen modaalilogiikalla, erit. episteeminen ja doksastinen logiikka.

Ma 4.4. Protokollat: Johdatusta sosiaaliseen valintaan ja mekanismien suunnitteluun.

Ti 5.4. Harjoitukset 3 (luennoista 21.3. ja 31.3.)

To 7.4. Oppiminen: Johdatusta monen agentin palauteoppimiseen.

Ma 11.4. Ongelmia ja haasteita: Peliteorian ja rationaalisuuden ongelmia ja rajoituksia, vaihtoehtoisia lähestymistapoja, tulevaisuuden näkymiä?

Ti 12.4. Harjoitukset 4 (luennoista 4.4. ja 7.4.)

Ti 19.4. Harjoitukset 5 (luennosta 11.4.)

Ti 26.4. Koe

Ke 11.5. Viimeinen tapaaminen, päivämäärä muuttunee ?