Department of Computer Science
Series of Publications C
Report C-2002-54

Klusterointimenetelmät-seminaari

Hannu Toivonen (toim.)

University of Helsinki
Finland

Klusterointimenetelmät-seminaari

Hannu Toivonen (toim.)

Tietojenkäsittelytieteen laitos
PL 26, 00014 Helsingin yliopisto

Raportti C–2002–54, Marraskuu 2002, vi+157 sivua

Aiheluokat (Computing Reviews 1998):
I.5.3 Clustering
H.2.8 Database Applications (Data Mining)
I.2.6 Learning

Sisällysluettelo

Esipuhe

Havaintojen klusterointi on niiden automaattista ryhmittelyä keskenään samankaltaisten havaintojen joukkoihin. Klusteroinnilla voidaan etsiä vastauksia esimerkiksi seuraaviin kysymyksiin:

- “Millaisiin luonteviin ryhmiin aineistoni voidaan jakaa?” (etsi klusterit)
- “Millaiset havainnot ovat tyypillisiä tai edustavia?” (anna tyypilliset eri klustereiden edustajat)
- “Miten aineistoani voi luonnehtia lyhyesti?” (kuvaa klusterit koko aineiston asemesta)

Klusterointi on yksi tiedon louhinnan (data mining, knowledge discovery) perusongelmatyypeistä ja jo vanha tilastollinen ja koneoppimisongelma.

Tämä teos sisältää Helsingin yliopiston tietojenkäsittelytieteen laitoksen syksyn 2002 klusterointiseminaarin raportit (58302306 Klusterointimenetelmät, 2 ov). Seminaari on luonteeltaan opintoseminaari, ja siinä käsitellään klusterointialgoritmeja sekä muutamaa sovellusta alkuperäisartikkeleihin perustuen. Kokonaisuutena raportit antavat kattavan yleiskuvan klusterointimenetelmistä — sekä jo perinteisistä että uusimmista menetelmistä.

Seminaarin toiminta matkii tieteellisiä konferensseja. Kukin osallistuja laati aiheestaan alkusyksyn aikana raportin (“artikkelin”), joka palautettiin seminaarin (“konferenssin”) järjestäjälle. Kunkin raportin arvioi kirjallisesti kolme muuta seminaarin osallistujaa (“ohjelmatoimikunta”). Palautteen perusteella tekijät muokkasivat raportit lopulliseen muotoonsa. Seminaari (“konferenssi”) kokoontuu lopulta kaksipäiväisenä, ja kukin osallistuja pitää esitelmän omasta raportistaan.

Seminaarin viimeistellyt raportit on koottu yhteen tähän “konferenssijulkaisuun”. Ansio tästä julkaisusta kuuluu seminaarin osallistujille: jo raporttikäsikirjoitukset olivat korkeatasoisia, ja “ohjelmatoimikunta” teki loistavaa työtä antaessaan perusteellista ja harkittua palautetta, jonka avulla osallistujilla oli mahdollista vielä parantaa ja viimeistellä raporttejaan tähän kokoelmaan.

Kiitos kaikille seminaarin osallistujille! Toivotan myös antoisaa “konferenssia” osallistujille sekä mielenkiintoisia lukuhetkiä muille tämän kokoelman lukijoille.

Helsingissä marraskuussa 2002

Hannu Toivonen