



Aalto-yliopisto
Perustieteiden
korkeakoulu

Perustieteiden korkeakoulu
<http://sci.aalto.fi/>
puhelin 050-365 2652 /Laura Öhrnberg

Väitöstiedote

26.10.2011

Aineiston satunnaistaminen paljastaa tiedonlouhinnan tulosten merkitsevyyden

Väitöskirjan nimi	Randomization Algorithms for Assessing the Significance of Data Mining Results Satunnaistusalgoritmeja tiedonlouhinnan tulosten merkitsevyyden arviointiin
Väitöskirjan sisältö	<p>Tiedonlouhinta on monitieteellinen tutkimusalue, joka kehittää yleisiä menetelmiä mielenkiintoisen ja hyödyllisen tiedon löytämiseksi suurista tietoaaineistoista. Tiedonlouhinnan avulla voidaan esimerkiksi parantaa dokumenttihaun onnistumista tai tunnistaa taudeissa samalla tavalla vaikuttavia geenejä mitattujen geeniaktiivisuuksien avulla. Suurten aineistojen kanssa työskennellessä ongelmana on, että menetelmät voivat löytää täysin satunnaisestakin aineistosta järkevältä näyttäviä tuloksia, jotka ovat kuitenkin pelkkää satunnaiskohinaa. Tässä väitöskirjassa esitellään uusia menetelmiä ja algoritmeja tulosten luotettavuuden arviointiin. Uusien menetelmien avulla voidaan tunnistaa, ovatko saadut tiedonlouhinnan tulokset vain satunnaisilmiöitä aineistossa vai jotakin mielenkiintoisempaa.</p> <p>Väitöskirjassa kehitetään ja sovelletaan satunnaistukseen perustuvaan merkitsevyydestausta, jossa alkuperäistä tulosta verrataan satunnaistetuilla aineistoilla saatuihin tuloksiin. Satunnaistetuilla aineistoilla ja alkuperäisellä aineistolla oletetaan olevan joitain yhteisiä perusominaisuuksia. Alkuperäisellä aineistolla saadut tulokset, jotka poikkeavat selvästi satunnaistetuilla aineistoilla saaduista tuloksista, eivät selity suoraan tutkituilla perusominaisuuksilla. Väitöskirjassa kehitetään uusia satunnaistusmenetelmiä neljään erityiseen laskennalliseen tilanteeseen. Menetelmät toimivat hyvin erityyppisillä aineistoilla, ovat helppokäyttöisiä ja antavat mielekästä tietoa tiedonlouhinnan tulosten parantamiseen ja ymmärtämiseen. Tulokset osoittavat, että kehittyneet satunnaistusalgoritmit ja niihin pohjautuva merkitsevyydestaus on hyödyllinen työkalu tulosten luotettavuuden arviointiin erilaisissa käytännön sovelluksissa.</p>
Väitöskirjan ala	Informaatiotekniikka, tiedonlouhinta
Väittelijä	Markus Ojala, DI Syntynyt Porissa 1984 Kirjoittanut ylioppilaaksi Porin suomalaisen yhteislyseon lukiosta 2003
Väitöksen ajankohta	12.11.2011 klo 12
Paikka	Aalto-yliopiston perustieteiden korkeakoulun E-sali, Otakaari 1, Espoo
Vastaväittäjä	Professori Rosa Meo, University of Torino, Italia
Valvoja	Professori Heikki Mannila, Aalto-yliopisto, Tietojenkäsittelytieteen laitos
Väitöskirjan verkko-osoite	http://lib.tkk.fi/Diss
Väittelijän yhteystiedot	Markus Ojala, Aalto-yliopisto, Tietojenkäsittelytieteen laitos, PL 15400, 00076 Aalto, p. 050 412 4884, Markus.Ojala@aalto.fi

Väitöskirja on julkisesti nähtävillä Aalto-yliopiston perustieteiden korkeakoulun ilmoitustaululla Konemiehentie 2, Espoo.