

20.05.2005 Asiakastapaaminen

-Olisi hyvä jos dokumentaatio/koodi olisi englanniksi.

-ohjelma pähkinänkuoressa:

-syöte: mallin kuvaus, todennäköisyysmalli, muuttujien kombinaatioille todennäköisyydet, (bayes-verkko?)

-syöte: aineisto, joidenkin mallin muuttujien arvoja.

-Ohjelma laskee kuinka jakauma muuttuu annetuilla arvoilla, kun arvot kiinnitetään muuttujille.

-tuloste: otoksia posteriorijakaumasta, aineistolla ehdollistetusta. Lukuja, tuntemattomien muuttujien arvoja. Kun näitä tulostetaan paljon, saadaan tilastollisia suureita, jakauman varianssia ja odotusarvoa, etc.

-ehdotusjakauma on tuntemattomille muuttujille ehdotettu jakauma, tästä otetaan tai hylätään tietyllä valinnalla.

-graafinen käli: mallien kuvaamiseen?, tämä voidaan tehdä tekstitiedostona, mutta usein muuttujien joukot yhdistyvät tiettyyn kokonaisuuteen (ts. Järven lämpötila, monta järveä, olisi kiva jos tällaisen toistorakenteen voisi esittää, graafisella käyttöliittymällä tai tekstitiedostolla. Käli ei ole prioriteetilla Primary, ainakaan.

-onko mallilla standardia? Olemassa bassistin vanhat kuvauskielet. Tässä ei ole tarkoitusta tehdä mitään niin yleistä. Rajoitutaan pienemmälle alueelle. Voidaan käyttää jotakin olemassa olevaa mallia, mutta ei samaa kuvauskieltä sellaisenaan.

-minimitavoite: Muutaman perusmallin tehokas simulointi, jopa ilman kuvauskieltä...

-olisi kiva jos vähän yleisemmällä tasolla, myös.

-rinnakkaisuus?: Ei korkealla prioriteetilla, voisi olla `liikaa`. Ratkaisut eivät tule olemaan sellaisia, että niitä joka kohdassa saa käyttöön rinnakkaisuutta. Olisi kiva toteuttaa. (priority: Tertiary)

Erityisvaatimuksia: linux riittää, windowsia ei tarvita.

-Aikaisemmasta versiosta: bassist koekäyttöön?,

-miksi ollaan tekemässä: Mahdollista analysoida paikkatietoa sisältäviä aineistoja, joka tuo uusia riippuvuuksia (vierekkäiset ruudut korreloivat..) Bassistilla tällaisten mallien teko hankalaa, ja bassist varsin tehoton. Esim. Suomi jaettuna 10x10km ruutuihin, menee vuosi... Bassistissa ei teknisiä säätöjä, joilla saa jotakin aikaan.

Ajoja kokeillaan eri vaihtoehtoilla, usein ikävää jos yhdessä ajossa menee kaksi viikkoa.

Hankalaa löytää yleisen tason kuvauskieli, jolla mahdollista esittää kaikki...

Esim. Funktioiden kohdistaminen tiettyihin muuttujiin ja niiden keskiarvojen mm. Laskeminen

...

-Ei kannata tehdä kovin yleistä rakennetta.

Löydetään kuvaustaparunko, johon voi lisätä uutta kamaa mahd. Helposti?

USE moduuleja?

-kuinka isoja, työlästä bassist-tekstitiedoston teko? - Voi joutua käyttämään skriptiä, joka generoi mallin kuvauskieltä...

-spatiaalisia riippuvuuksia: markovin satunnaiskenttiä. Ohjelmassa voitava kytkeä tämä pois. Ideana vertailla spatiaalista riippuvuutta sen puuttumiseen, tuleeko erilaisia tuloksia?

Tulosten käsittelystä: Olisi hyvä että voisi laskea tulostejakaumasta keskiarvon luottamusvälin rajat. (tai bayesiläisesti jotain muuta, tässä käytetään bayesilaista päättelyä, huom.)

Komponentteja outlined:

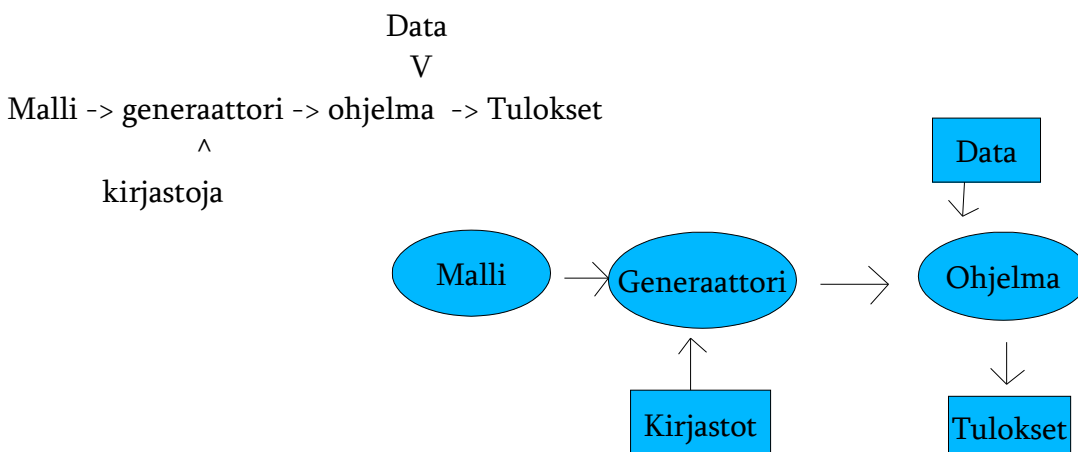
-käli: pystyy auttamaan mallien kirjoittamisessa,

-mallitulkki: ymmärtää tietynlaisia riippuvuusmalleja ja kertoo ohjelmapalikalle minkäläistä koodia halutaan,

-tulkin tulosten suoritettavaksi ohjelmaksi muuttava palikka: ottaa tulkin tulostetta vastaan, kirjoittaa koodin, kääntää sen.

-Matemaattista pohjaa esitellään myöhemmin.

-todennäköisyyslaskennan perusteet, etc.



Malliin voi liittyä datan kuvaus ja kertoa mihin data kuuluu.

-priorijakaumat halutaan antaa, halutaan antaa hyväksymistodennäköisyys muutoksille, etc..

-nämä annettava generaattorille.

-bassistissa oli implementaatio-osuus tekstitiedostossa.

-Tärkeää, että tämä onnistuu.

Generaattori tuottaa fortran-koodia, joka laskee posteriorijakaumat.

Olemmeko tekemässä bassistia uudestaan? Onko tässä järkeä?

Kannattaisiko parantaa bassistia, ottaa sen kielestä osajoukko ja muuttaa sen generaattori tekemään tehokasta fortrania? Malliin liitettävä linkitetty lista.

Malleista puuttuu symmetrisyys.

-Projektia rajattava tarkasti, jotta tästä tulee jotain järkevää.

-Protoilua? Lähdetään yhdestä malliesimerkistä, tehdään siihen ratkaisu, laajennetaan?

-Tehdään pieni ydin, laajennetaan.

Rakenne saatava kunnolliseksi, jotta sitä voidaan laajentaa.

Generaattorin suunnittelu: minkälainen on minimaalinen malli, joka pystyy esittämään esimerkkidatat? Kuinka paljon simulaatioteknisiä parametreja pitää muuttaa?

-tehdään mallidatalle käsin (lopullisessa generaattorin tekemä) sen laskeva ohjelma parille datalle eri, katsotaan yhteyksiä, auttaa generaattorin suunnittelussa.

-Generaattori vaatii kunnollista arkkitehtuurisuunnittelua, että sen voi tehdä.

-Seuraavaksi mallia ja dataa, esimerkkiä...

249 lintulajia, yhtä analysoidaan 3813 ruudussa, ruutuihin liittyy param, lajiin liittyy kussakin ruudussa pesimävarmuusindeksi (0-3, int 3 varmistettu pesintä, muut alemmas). Jokaiseen ruutuun liittyy havainnointiaste (kok. Luku 0-5, käytännössä 1-5, int). Jokaiseen ruutuun liittyy myös selittäviä muuttujia: Kasvillisuusvyöhykejako, etc. Ruutuihin liittyy vierekkäisyysuhde, jolla ne riippuvat toisistaan mm. Pesintä.

20 000 sanaa, näihin liittyy joukko kuntia, joissa näytä tiedetään käytettävän.

1 2 3 4 ..

1 0 1 0 1

2 1 1 0 1           etc, joka sanaan liittyy rivi kuntia joissa se esiintyy,

3                    kuin fortran-harjoitustyössä

4

Paikkatietodataa: Suomen peruskartan paikannimet, tarkat koordinaatit. Tästä voitaisiin tehdä murresanadataa, kunnan alueelta kaikki kota-sanat sisältävät paikannimet, tai jotain...