

# NIP

## Mikä ihme se on?

Kuvaus diskreettien muuttujien inferenssistä  
aikasarjoissa  
Janne Toivola [jatoivol@niksula.hut.fi](mailto:jatoivol@niksula.hut.fi)

## Taustaa

- NIP-nimen historia (vielä ehtii antaa uuden nimen)
- Saman kaltaisia ohjelmia tehty muuallakin
- Tanskalaisten Hugin Expert
- Adnan Darwiche & co: SAMIAM
- Ei tarkoitettuja aikasarja-analyysiin

## Mitä iloa päättelystä

- Bayesilaisella päättelyllä löydetään piilomuuttujien tn-jakaumat...
- Otetaan huomioon esim. aika, paikka, positio kromosomissa tms. konteksti
- Luokittelu, segmentointi jne. piilomuuttujien avulla
- Automaattinen puutteellisen datan käsittely

## Teoriaa: verkko

- Pohjana muuttujien riippuvuudet ilmaiseva suunnattu sykliton verkko DAG
- Mallin parametrit muotoa  $p(c|pa(c))$
- C on riippumaton muista kuin vanhemmistaan  $pa(C)$
- Muuttujia kuvaavien solmujen välillä kaaret vanhemmista lapsiin

## Teoriaa: liittymäpuu

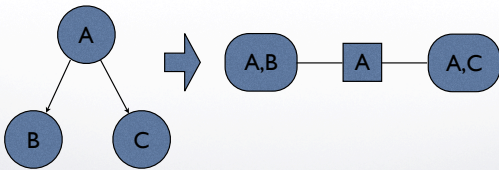
- E.m. verkko soveltuu huonosti inferenssin vaatimiin laskutoimituksiin
- Täysi yhteis-tn-jakauma tuhlaa muistia
- Käytettävä vaihtoehtoista esitystapaa nimeltä liittymäpuu
- Puun solmut tallentavat nk. "perheiden" yhteistodennäköisyysjakaumia

## Teoriaa: liittymäpuu

- Puuhun tallennetut jakaumat nimeltään potentiaaleja (mahd. normalisoimaton)
- Minimaalinen tapa esittää yhteis-tn-jakauma pienempien tulona
- Vrt.  $p(x, y, z) = p(x|y) * p(y|z) * p(z)$
- Myös ns. *running intersection property*

## Teoriaa: liittymäpuu

- Esimerkki graafista ja liittymäpuusta



## Teoriaa: päättely

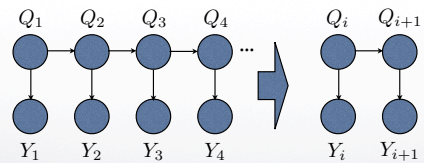
- Päättely tapahtuu “levittämällä” potentiaaleja painottavia jakaumia naapurisolmuihin puussa
- Marginalisoimalla saadaan pienempi projekzio potentiaalista:  $\phi(A) = \sum_B \phi(A, B)$
- Kertolaskulla päivitetään potentiaalia uuden evidenssin mukaan:  $\phi(A, C) := \phi(A, C) * \phi(A)$

## Teoriaa: aikaviipaleet

- Aikasarjoja varten toistuvakenteiset mallit
- Esim. HMM toistaa pientä “verkkoa”
- Esitetään malli viipaleena, jota toistetaan
- Määritettävä toistuva osa liittymäpuusta laskutoimituksia varten
- Viipaleen käyttö päättelyssä “puoliaskelin”

## Teoriaa: aikaviipaleet

- Esimerkki aikaviipaleesta: HMM



## Toteutus: yleistä

- C-kielinen ohjelmakirjasto
- Huginin net-tiedostojen jäsenys
- Muunnos graafeista liittymäpuuksi
- Aikasarjan luku tiedostosta
- Edellä kuvattu inferenssi mahd. aikaviipalein
- EM-algoritmi parametrien viilaamiseen

## Hugin net-formaatti

- Malli luetaan Huginin asettaman “standardin” mukaisesta tiedostosta
- node-määrittelyt kertovat muuttujat
- Oma *NIP\_next* kenttä aikaviipalemalleille
- potential-määrittelyt ehdollisille todennäköisyysjakaumille
- Graafin rakenne samassa implisiittisesti

## Graafista liittymäpuuksi

- By Antti Rasinen, kiitokset
  - Muunnos heti net-tiedoston jäsenyyksen jälkeen
  - Alkuperäinen graafi unohdetaan
  - malli = liittymäpuu + muu kirjanpito
- ```
Nip model = parse_model("hmm.net");
```

## Datan luku tiedostosta

- Tiedoston 1. rivillä muuttujien nimet
- Loput rivit aikasarjan askelia
- Kullakin rivillä muuttujien arvojen *nimet*
- Luetaan data tiedostosta aikasarjaa esittäväksi tietorakenteeksi

```
TimeSeries ts = read_timeseries(model,"data.txt");
```

## Päättely

- Inferenssialgoritmi antaa esim. "epävarman aikasarjan"
- Annettujen muuttujien kunkin arvon todennäköisyys kullakin hetkellä
- Esim. yhteistodennäköisyydetkin mahdollisia, mutta implementoimatta

```
UncertainSeries ucs =  
forward_inference(ts, ts->hidden,  
ts->num_of_hidden);
```

## EM-algoritmi

- Toteutus vielä kesken
- Algoritmille annetaan opetettava malli alkuarvauksena, aikasarjallinen dataa ja kynnsarvo lopetusehtoa varten
- Aikasarja sisältää viittauksen malliin

```
em_learn(timeseries, threshold);
```

## Vielä muuta?

- Myös ns. "pehmeää evidenssiä" voisi syöttää havaintoina
- Parametrien kirjoittaminen net-tiedostoon
- Keinotekoisien datan generointi annetun mallin perusteella
- Paljon sovelluksia ja dataa?