Kehittynyt Päättely Bayes-verkoissa

Robert Cowellin kappaleesta *Advanced Inference in Bayesian Networks* Michael I. Jordanin (toim.) kirjassa *Learning in Graphical Models*

Mika Inki

3. Lokakuuta, 2001

Aluksi

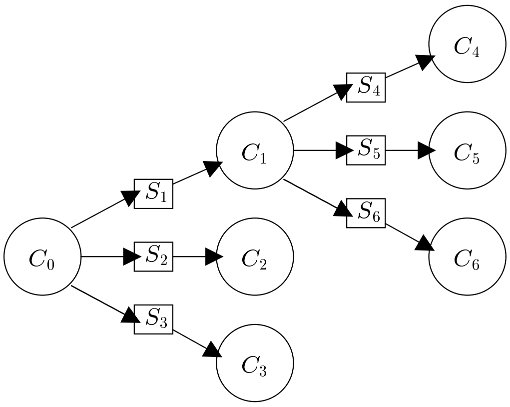
- Jatkuu kutakuinkin edellisestä kappaleesta (sama kirjoittaja, Robert Cowell). Käytetään liittymäpuita
- Käsitellään:
  - Otantaa
  - Uskottavimpia konfiguraatioita
  - Fast retraction (“nopea takaisinveto”)
  - gaussisia ja ehdollisesti gaussisia malleja
- Melko teknistä luonteeltaan...

Otanta (diskreetissä tapauksessa)

- DAG:n (Suunnattu syklitön verkko) solmut voidaan pistää topologiseen järjestykseen:
  $(X_1, X_2, \ldots, X_n)$, missä X_i:t ovat verkon solmut, ja solmun i vanhemmat ovat ennen itse solmua
  (eli vanhemman indeksi $j < i$)
- Datapisteen luominen kun ei ole havaintoja:
  generoidaan ensin X_1:n arvo x_1^* jakaumasta $P(X_1)$, sitten arvot x_j^* jakaumasta $P(X_j|\mathrm{pa}(X_j) = \pi^*)$
  Lopulta meillä on arvo $u^* = (x_1^*, \ldots, x_n^*)$
- kun on havaintoja, ε:
  Sama kuin edellä mutta ellei $x_j^* = x_j^\dagger \quad \forall j$, joilla $x_j^\dagger \in \varepsilon$, hylätään nykyinen yrite ja aloitetaan X_1:stä uudelleen.
  (ei voida suoraan asettaa $x_j^* = x_j^\dagger$, koska tällöin arvo ei olisi oikeasta jakaumasta)
- Hylkäysmäärä kasvaa eksponentiaalisesti havaintojen lukumäärässä - ehkä liittymäpuita voisi käyttää apuna.

Otanta käyttäen liittymäpuuta

- Oletetaan havaintoja ja seuraava esitys yhteisjakaumalle:
  $P(U|\varepsilon) = \prod_C P(C|\varepsilon) / \prod_S P(S|\varepsilon)$
- Olkoon joku (liittymäpuun) klikki C_0 asetettu juureksi. Tehdään klikkien ja separaattoreiden välisistä reunoista suunnattuja s.e. ne osoittavat poispäin juuresta. (Kyseessä on puu -> onnistuu.)
  Saadaan topologinen järjestys: $(C_0, S_1, C_1, \ldots, S_m, C_m)$
- Voidaan muodostaa esitys (set-chain: joukko-ketju):
  $P(U|\varepsilon) = P(X_{C_0}|\varepsilon) \prod_{i=1}^m p(X_{C_i \setminus S_i}|X_{S_i}, \varepsilon)$
- Ota $x_{C_0}^*$ jakaumasta $P(X_{C_0}|\varepsilon)$. Seuraavaksi $x_{C_1}^*$ jakaumasta $P(X_{C_1 \setminus S_1}|X_{S_1}, \varepsilon)$, jne. Ei tarvita hylkäystä.
  (Tosin hylkäystä ei tarvita koska käytetään jakaumia ehdolla havainnot - liittymäpuilla ei ole tekemistä tämän kanssa!)



Kuva 1: Suunnattu liittymäpuu

Maksimin kuljetus ja todennäköisimmät konfiguraatiot

- Oletetaan $P(U,\varepsilon) = \prod_C a(C) / \prod_S b(S)$ ja viesti $b^*(S) = \max_{C \setminus S} a(C)$
- saadaan: $P(U,\varepsilon) = \frac{\prod_C P^{max}(C,\varepsilon)}{\prod_S P^{max}(S,\varepsilon)} = \frac{\prod_C \max_{U \setminus C} P(U,\varepsilon)}{\prod_S \max_{U \setminus S} P(U,\varepsilon)}$
- Esim: oletetaan klikit A, B ja B, C, sekä viesti $b^*(B) = \max_C a(B,C)$
- $P^{max}(A,B,\varepsilon) = \max_C P(A,B,C,\varepsilon) =$
  $\max_C a(A,B)\frac{1}{b(B)}a(B,C) =$
  $(= \max_C a(A,B)\frac{b^*(B)}{b(B)}\frac{1}{b^*(B)}a(B,C))$
  $a(A,B)\frac{b^*(B)}{b(B)}$
  (Tämä riippuu nyt vain juuriklikistä.)
- Voidaan kuljettaa maksimit juureen (hakien juuren maksimin), ja ottaa näin saatu arvo lisähavainnoksi, jonka avulla etsitään seuraavan tason solmujen arvot, jotka otetaan myös lisähavainnoiksi, jne. Lopulta saadaan uskottavin konfiguraatio.

Maksimin rappeutuneisuus

- halutaan tietää montako erilaista muuttujakonfiguraatiota antaa maksimin.
- olkoon $P^{max}(U|\varepsilon) = p^*$

$$I_C(X_C|\varepsilon) = \begin{cases} 1 & \text{jos } P_C^{max}(X_C|\varepsilon) = p^* \\ 0 & \text{muuten} \end{cases}$$

$$I_S(X_S|\varepsilon) = \begin{cases} 1 & \text{jos } P_S^{max}(X_S|\varepsilon) = p^* \\ 0 & \text{muuten} \end{cases}$$

- $I(U|\varepsilon) = \frac{\prod_C I_C(X_C|\varepsilon)}{\prod_S I_S(X_S|\varepsilon)}$ on nyt indikaattorifunktio uskottavimmalle konfiguraatiolle.
- kuljetetaan summaa -> saadaan maksimin kertaluku ("oikean elämän" tapauksissa kertaluku on yleensä 1)

N parasta konfiguraatiota

- Olkoon $(X_1, \ldots, X_N)$ joku järjestys, ja olkoon i:nnellä solmulla k_i eri tilaa $(x_1^{k_1}, \ldots, x_n^{k_n})$
- oletetaan että ei ole havaintoja, ja olkoon $M^1 = (x_1^{m_1^1}, \ldots, x_N^{m_1^1})$ uskottavin konfiguraatio
- toiseksi uskottavin konfiguraatio M^2 eroaa ainakin yhdessä kohtaa M^1:stä
- otetaan pseudohavaintoja $\varepsilon_1^1 = \{X_1 \neq x_1^{m_1^1}\}$
  $\varepsilon_2^1 = \{X_1 = x_1^{m_1^1} \quad \wedge \quad X_2 \neq X_2^{m_1^1}\}$ jne.
  ε_n^1:een saakka. Haetaan maksimi kullakin näistä havainnoista. Jollakin ε_j^1 saadaan M^2.
- M^3 löytyy joko joukosta $\{\varepsilon_l^1 | l \neq j\}$ tai se eroaa M^2:sta jollain arvolla $j_2 \geq j$. Tämä j_2 voidaan hakea samalla tavalla käyttäen havaintoja $\varepsilon_j^2 = \{\varepsilon_{j-1}^1 \quad \wedge \quad X_j \neq x_j^{m_j^2} \wedge X_j \neq x_j^{m_j^1}\}$ jne.
- jos on etukäteishavaintoja, otetaan ne huomioon (vähentää tarvittavien hakujen määrää).

L^p kuljetus

- Minimiä ja summaa voidaan kuljettaa samalla tavalla kuin maksimia.

- Kaikkien näiden voidaan ajatella olevan erikoistapauksia L^p-normin kuljetuksesta.

- $L^p(f) = (\int_{x\in X} f^p(x)dx)^{\frac{1}{p}}$
maximi: $p \to \infty$,
minimi: $p \to -\infty$
summa: $p = 1$

- Viesti: $b^*(X_S) = (\sum_{X_C \setminus X_S} a_C^p(X_C))^{\frac{1}{p}}$
ja $P(u|\varepsilon) = \frac{\prod_C P_C^{L^p}(X_C|\varepsilon)}{\prod_S P_S^{L^p}(X_S|\varepsilon)}$

"Nopea takaisinveto"

- Joskus halutaan tutkia myös havaintojen todennäköisyyttä: Matala tn. kertoo jos mallissa on ongelmia tai jos kyseessä on harvinainen tapaus.

- voitaisiin poistaa yksi havainto u vuorollaan havainnoista ε, eli $\varepsilon_{\setminus u}$, ja sitten hakea tn. poistetulle havainnolle.

- voidaan kuitenkin käyttää nopeampaa algoritmia, jota kutsutaan nimellä fast-retraction ja joka hyödyntää X:ään liittyvien havaintojen ε_X uskottavuutta h_X.

- käytetään "ulosmarginalisointia" (out-marginalisation): otetaan kaksi klikkiä A, B ja B, C.
Yhteisjakauma: $P(ABC, \varepsilon_A) = P(ABC)h_A$
Viesti: $g^*(B) = \sum_C g(BC)h_C$

- $P^{out}(AB, \varepsilon_{\setminus A\cup B}) = \sum_C P(ABC, \varepsilon_C) =$
$= \frac{g(AB)g^*(B)}{g(B)}$
$P(U) = \frac{\prod_C P^{out}(C, \varepsilon_{\setminus C})}{\prod_S P^{out}(S, \varepsilon_{\setminus S})}$

Mallinnus käyttäen jatkuvia muuttujia

- Tähän asti on käsitelty vain diskreettejä muuttujia. Voidaan kuitenkin käyttää jatkuviakin muuttujia.

- Integrointeja ei kuitenkaan voida yleisesti laskea (analyyttisesti) käytettäessä jatkuvia muuttujia. Täytyy käyttää simulaatiomenetelmiä, esim. Gibbsin otantaa. (Raskas.)

- Jos kuitenkin kokonaisjakauma on gaussinen, ovat integraalit laskettavissa. Tätä voidaan laajentaa tapaukseen, jossa on sekä jatkuvia että diskreettejä muuttujia, eli ehdollisesti gaussisiin malleihin.

- Esimerkki gaussisesta mallista:

$$\boxed{Y} \to \boxed{X} \to \boxed{Z}$$

$y \sim N(\mu_y, \sigma_y^2) \qquad (= P(Y))$
$x \sim N(\mu_x + \beta_{x,y}y, \sigma_x^2) \qquad (= P(X|Y))$

$z \sim N(\mu_z + \beta_{z,x}x, \sigma_y^2) \qquad (= P(Z|X))$
Yleisesti, jos x:llä on vanhemmat y_i, $\quad i = 1, \ldots, n$, niin $x \sim N(\mu_x + \sum_i \beta_{x,y_i}y_i; \sigma_x^2)$

- $P(X,Y,Z) = P(Y)P(X|Y)P(Z|X) =$
$N([\mu_x\mu_y\mu_z]^T, K) \quad \propto \quad \exp(\bar{x}^T h - \frac{1}{2}\bar{x}K\bar{x})$

- Gaussiset potentiaalit:
Osajoukon $\{Y_1, \ldots, Y_k\}$ ($\subset \{X_1, \ldots, X_n\}$) potentiaali $\phi(g,h,K)$ on $\exp(g+\bar{y}^t h_y - \frac{1}{2}\bar{y}^T K_y \bar{y})$, missä $\bar{y}$ sisältää muuttujat y_i täydennettynä nollilla. (Samaten K_y.)

- potentiaalilla on yksinkertaiset laskutoimitukset:
$\phi_1 * \phi_2 = \phi(g_1 + g_2, h_1 + h_2, K_1 + K_2)$
$\phi_1 / \phi_2 = \phi(g_1 - g_2, h_1 - h_2, K_1 - K_2)$

- Marginalisoinnissa integroidaan muuttuja pois (yksinkertainen algebrallinen muoto).

- liittymäpuiden rakenne gaussisissa verkoissa on sama kuin diskreetissä tapauksessa

- jokaiselle klikille luodaan oikeankokoinen gaussinen potentiaali

- Yhteisjakaumalla sama esitys kuin diskreetissä tapauksessa:
  $P(u) = \frac{\prod_C P(X_C)}{\prod_S P(X_S)}$,
  missä P:t ovat gaussisia potentiaaleja.

- havainnot ovat täsmällisiä arvoja (ei vain poissulkevia), ja havainnot täytyy laittaa sisään kaikkiin klikkeihin, joissä ko. muuttuja esiintyy

Ehdollisesti gaussiset mallit

- Sekoitus jatkuvista ja diskreeteistä muuttujista

- diskreeteillä muuttujilla ei saa olla jatkuvia vanhempia

- diskreettien muuttujien todennäköisyydet ovat taulukoita, jatkuvien muuttujien gaussisia potentiaaleja, jossa kolmikot $\{g, h, K\}$ (kanoniset karakteristikat) on taulukoitu diskreettien vanhempien konfiguraatioiden mukaan. (Lisäksi jotkut jotkut konfiguraatiot saattavat olla kiellettyjä)

- yhteisjakauma on muotoa $f(x) = f(i, y) = \chi(i)\exp[g(i) + \bar{y}^T h(i) - \bar{y}^T K(i)\bar{y}^T]$, missä $\chi(i) = 1$, jos konfikuraatio i on sallittu, muuten 0. Potentiaali on samaa muotoa.

- kan. karakt. ovat määriteltyjä vain kun $\chi > 0$, kuten myös momenttikarakteristikat $\{p, \xi, \Sigma\}$ (tn., keskiarvo, varianssi)
  $\xi(i) = K(i)^{-1}h_i, \qquad \Sigma(i) = K(i)^{-1}$

Vahva ja heikko marginalisointi

- Jatkuvien muutujien marginalisointi integroinnilla; saadaan (edelleen) ehdollisesti gaussinen potentiaali (CG-potentiaali).

- Diskreettien muuttujien marginalisointi vastaa summaamista; saadaan sekoitus CG-potentiaaleista.

- Diskreettien muuttujien yli marginalisoinnin tulos määritellään CG-potentiaaliksi, jonka momenttistatistiikat $\{\tilde{p}, \tilde{\xi}, \tilde{\Sigma}\}$ ovat:
  $\tilde{p}(i) = \sum_j p(i,j)$
  $\tilde{\xi}(i) = \sum_j \xi(i,j)p(i,j)/\tilde{p}(i)$
  $\tilde{\Sigma}(i) = \sum_j \Sigma(i,j)p(i,j)/\tilde{p}(i) +$
  $+ \sum_j (\xi(i,j) - \tilde{\xi}(i))^T(\xi(i,j) - \tilde{\xi(i)})p(i,j)/\tilde{p}(i)$

- Marginalisointi molempien muuttujatyyppien yli: Ensin marginalisoi jatkuvien muuttujien yli, ja sitten diskreettien muuttujien yli.

- Mikäli jatkuvien muuttujien yli marginalisoinnin tuottama pari (h, K) on riippumaton niistä diskreeteistä

muuttujista, joiden yli tullaan marginalisoimaan, on kyseessä vahva marginalisointi, muuten täytyy summata diskreettien muuttujien yli ja kyseessä on heikko marginalisointi.

- sekä heikko ja vahva marginalisointi täyttävät ehdon
  $\sum_A (\sum_B \phi_{A\cup B\cup C}) = \sum_{A\cup B} \phi_{A\cup B\cup C}$
  mutta (yleisesti) vain vahva ehdon
  $\sum_A \phi_{A\cup B}\psi_B = \psi_B \sum_A \psi_{A\cup B}$

- molemmilla marginalisointitavoilla saadaan 'marginalisoitu' jakauma, jolla on oikeat momentit kertalukuun 2 saakka.
  $P(I = i) = \tilde{p}(i)$, $E\{Y|I = i\} = \tilde{\xi}(i)$,
  $V\{Y|I = i\} = \tilde{\Sigma}(i)$

Liittymäpuiden käyttö CG-verkoissa

- Liittymäpuun muodostuksessa täytyy hakea vahva juuri:
  Vahva juuri on mikä tahansa klikki R, jolle mikä tahansa keskenään naapuriklikkipari A, B, joista A on lähempänä R:ää, pätee:
  $(B \setminus A)$ on jatkuvien muuttujien osajoukko tai $(B \cap A)$ on diskreettien muuttujien osajoukko.

- Kerätään viestit vahvaan juureen, jossa viestit muodostetaan vahvan marginalisoinnin avulla.

- Levitetään viestit käyttäen heikkoa marginalisointia.

Lopuksi

- Käsiteltiin päättelyä liittymäpuissa

- Ei oikeastaan muita johtoajatuksia ...

- Lisälukemista (aloittakaa kappaleen osiosta 10):

Fast Retraction of evidence in a probabilistic expert system, (Cowell et al., 1992)
Propagation of probabilities, means and variances in graphical association models, (Lauritzen, 1992)