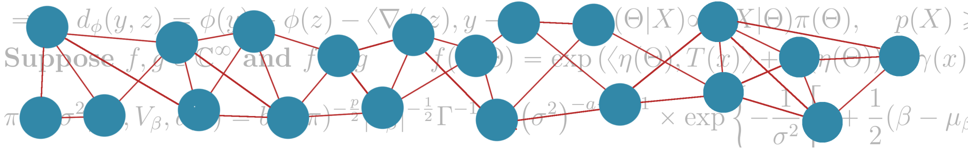


Distribuované dynamické bayesovské modelování

Kamil Dedecius

ÚTIA AV ČR, v.v.i.

27.3.2014

Malá motivace úvodem

Senzorové sítě

- Především WSNs (Wireless sensor networks)
- monitorování oblastí, detekce průniků, radar
- monitorování kvality vzduchu, vody aj.
- odhalování rizika povodní, radiace aj.
- průmyslové aplikace
- fyzikální experimenty

Hlavní výhody: cena, redundance, mobilita, relativní jednoduchost.

I chytrý telefon může být hloupým senzorem :-)

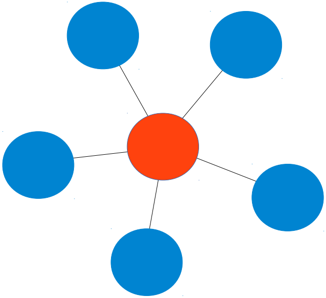
Topologie

Hvězda s fúzním centrem

- orientovaný graf (strom)
- vrcholy měří data
- FC zpracovává měření
- ...a případně posílá zpět

+ jednoduchost zpracování informace
+ porucha vrcholu se dá detekovat
+ solidní flexibilita (přidat/ubrat vrchol)

- SPoF v podobě FC
- vysoké komunikační nároky



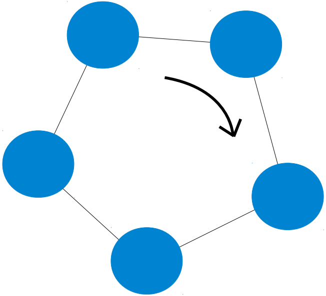
Topologie

Hamiltonovský graf ("token ring")

- orientovaný graf s jednou kružnicí všemi vrcholy
- žádná dedikovanost, žádné FC

+ jednoduchost zpracování informace
+ komunikačně nenáročné

- SPoF v každém uzlu
- porucha vrcholu se propaguje dál



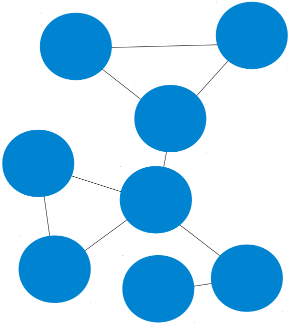
Topologie

Difuzní síť

- orientovaný graf, v krajním případě úplný
- žádná dedikovanost, žádné FC
- ostatní topologie jsou “speciální případy”

+ komunikačně “spíše” nenáročné
+ vysoká redundance (žádný SPoF)
+ značná flexibilita (přidat/ubrat vrchol)
+ porucha v uzlu se dá potlačit

- složitější zpracování informace



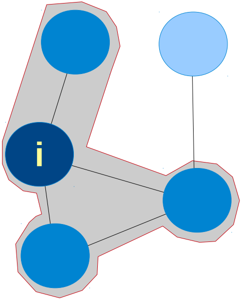
Jak taková spolupráce vypadá

Vezměme jako příklad difuzní síť

Cíl: k zadanému ($\sim$ společnému) modelu odhadnout společný param. Θ

Máme:

- $i = 1, \ldots, n$ – vrcholy
- $x_t^{(i)}$ – pozorování ve vrcholech
- $\mathcal{N}_i$ – množina sousedů vrcholu i (vč. i)



Vrchol i může s $j \in \mathcal{N}_i$ sdílet:

- pozorování $x_t^{(j)}, j \in \mathcal{N}_i$
- odhady Θ v $\mathcal{N}_i$
- (jedno, druhé, oboje)

Postupy fúze dat

Principy:

1. *difuze:* žádné meziiterace. Získat info – zapracovat – konec.
2. *koncensus:* meziiterace (optimalizace) – čas. a kom. náročné.

Difuzní kroky:

C – combine – fúze odhadů

A – adapt – fúze pozorování

Jsou-li A i C:

ATC – Adapt–Then–Combine

CTA – Combine–Then–Adapt

- Cattivelli, F.S. & A.H. Sayed, 2010. Diffusion LMS Strategies for Distributed Estimation.
- Tu, Sheng-Yuan & A.H. Sayed, 2012. Diffusion Strategies Outperform Consensus Strategies for Distributed Estimation Over Adaptive Networks.
- Mateos, G., I.D. Schizas & G.B. Giannakis, 2009. Distributed Recursive Least-Squares for Consensus-Based In-Network Adaptive Estimation.
- a mnohé další

Proč bayesovsky?

ACCP 37th Annual Meeting, Philadelphia, PA [2]

Differences Between Bayesians and Non-Bayesians
According to my friend Jeff Gill

Typical Bayesian

Typical Non-Bayesian

Protože bayesovci jsou sympaťáci :D (credit G. Casella)

Bayesovské modelování (*All models are wrong*)

$$\underbrace{\pi(\Theta|d)}_{\text{aposteriorno}} \propto \underbrace{p(d|\Theta)}_{\text{model}} \times \underbrace{\pi(\Theta)}_{\text{apriorno}}$$

Θ – latentní parametry modelu

- regresní koef., stř. hodnota, variance. . .

d – data (pozorovaní)

- na jejich základě odhadujeme Θ

$\propto$ – proporcionalita (vzorci chybí normalizace),

π – apriorní resp. aposteriorní h.p

- hyperparametry

Frekventismus

- fixní parametry
- pozorování (experimenty, studie) jsou opakovatelná
- odhady jsou bodové nebo intervalové (např. 95% IS)

Bayesovství

- fixní data (každé pozorování o něčem vypovídá)
- . . . ne nutně o tomtéž
- parametry jsou neznámá (leč fixní) veličina popsaná distribucí

Frekventismus

- 95% IS: 95% intervalů z opakovaného vzorkování (experimentů) pokrývá skutečný parametr.
- Bayes: mě nezajímá 95 ze 100 experimentů, já mám jen jeden!
- (jaká je tedy pst že v tvém IS parametr opravdu leží?)

Bayesovství

- 95% IS: dle dat pokrývá interval neznámý parametr s 95% pstí.
- Frekventista: a co když ve skutečnosti je pst jen 10%?
- (a také můžeš mít nesmyslné apriorno).

> *It does not help that many Bayesians over the years have muddied the waters by describing parameters (of models) as random rather then fixed. Actually, for Bayesians as much as any other statistician, parameters are fixed, but unknown. It is the knowledge about these unknowns that Bayesians model as random.*
> *(Gelman&Robert, AmStat 67(1), 2013)*

Bayesovství a dynamika

- Vzpomeňme

$$\pi(\Theta|d) \propto p(d|\Theta)\pi(\Theta)$$

- Dostaneme nové pozorování

$$d_2 \sim p(d_2|\Theta)$$

- Updatujme!

$$\pi(\Theta|d, d_2) \propto p(d_2|\Theta)\pi(\Theta|d)$$

Tedy ve vší obecnosti: Buďte $x_i, i = 1, \ldots, n$ pozorování náhodné veličiny $X \sim \mathcal{L}(X; \Theta)$ s h.p. $p(X|\Theta)$. Buď $\pi(\Theta)$ apriorní h.p. pro Θ. Potom

$$\pi(\Theta|x_1, \ldots, x_n) \propto \pi(\Theta) \prod_{i=1}^{n} p(x_i|\Theta)$$

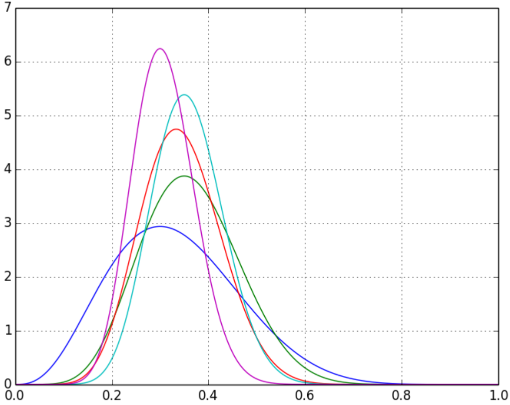
Notace odpustí...

Příklad

$X \sim \mathrm{Binom}(10, \mathrm{p})$, $p \sim \mathrm{Beta}(\mathrm{r}, \mathrm{s})$.

$r_0 = s_0 = 1$. skut. $p = 0.25$.

$\{x_1, \dots, x_5\} = \{3, 4, 3, 4, 1\}$.



Konjugovaná apriorna – exponenciální třída

$$p(X|\Theta) = \exp\left(\langle \eta(\Theta), T(x)\rangle - \Psi\left(\eta(\Theta)\right) + k(x)\right)$$

$\eta \equiv \eta(\Theta)$ – přirozený parametr ($\eta(\Theta) = \Theta \;\Rightarrow$ kanonická ET)

$T(x)$ – suficientní statistika (aditivita vzhl. k exp. fci)

Ψ – kumulant (log-partition) – normalizační fce

Má-li apriorno tvar

$$\pi(\Theta|\xi_{t-1}, \nu_{t-1}) = \exp\left(\langle \eta(\Theta), \xi_{t-1}\rangle - \nu_{t-1}\Psi\left(\eta(\Theta)\right) + l(\xi_{t-1}, \nu_{t-1})\right)$$

Je aposteriorno (pro n pozorování) dáno

$$\xi_t = \xi_{t-1} + T(x_1, \ldots, x_n), \qquad \nu_t = \nu_{t-1} + n.$$

Příklady: $\mathcal{N} \times \mathcal{N}$ (resp. $\mathcal{NG}, \mathcal{N}i\mathcal{G}, \mathcal{N}i\chi^2$), $\mathrm{Multi} \times \mathrm{Dir}$, $\mathrm{Binom} \times \mathrm{Beta}\ldots$

Motivace na úvod

Topologie

Jak spolupráce vypadá

Bayesovské principy

Míchání distribucí

Aplikace letem světem

Míchání distribucí

Kullbackova-Leiblerova divergence
Dány 2 h.p. p, q takové, že $q(x) \ll p(x)$, potom

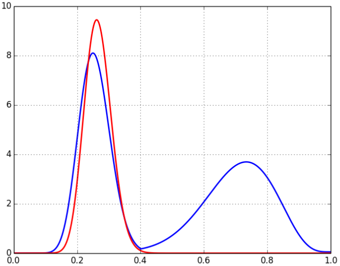
$$\mathcal{D}(p||q) = \mathbb{E}_{p(x)}\left[\log \frac{p(x)}{q(x)}\right] dx = \int p(x) \log \frac{p(x)}{q(x)} dx$$

$\mathcal{D}$ je premetrika (je nezáporná, nesymetrická a nesplňuje $\triangle$ nerovnost).

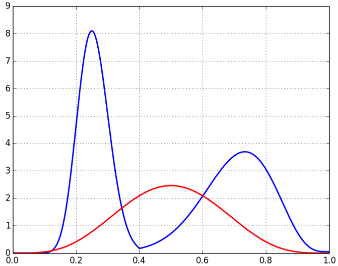
Jak se chová aproximace minimalizující KL?

Viz argumenty střední hodnoty...

Známé q (bimod), hledáme q (uni)



Známé p (bimod), hledáme q (uni)



Míchání distribucí

- Dáno: h.p. $f_1, \dots, f_n$ s nezáp. vahami $\omega_1, \dots, \omega_n$ se sumou 1
- Hledáme: h.p. q takovou, která je "nejblíže" ve smyslu

$$\sum \omega_i \mathcal{D}(f_i||q) \to \min \qquad \sum \omega_i \mathcal{D}(q||f_i) \to \min$$

$$q = \sum_{i=1}^{n} \omega_i f_i \qquad q \propto \prod_{i=1}^{n} f_i^{\omega_i}$$

??? Co vybrat ???

Odpověď: záleží na situaci. Podívejme se ještě jednou na KL…

Míchání distribucí

- Dáno: h.p. $f_1, \dots, f_n$ s nezáp. vahami $\omega_1, \dots, \omega_n$ se sumou 1
- Hledáme: h.p. q takovou, která je "nejblíže" ve smyslu

$$\sum \omega_i \mathcal{D}(f_i || q) \to \min \qquad \sum \omega_i \mathcal{D}(q || f_i) \to \min$$

$$q = \sum_{i=1}^{n} \omega_i f_i \qquad q \propto \prod_{i=1}^{n} f_i^{\omega_i}$$

??? Co vybrat ???

Odpověď: záleží na situaci. Podívejme se ještě jednou na KL...

- Jsou-li distribuce velmi podobné, bude rozdíl (dostatečně) malý.
- Jsou-li distribuce velmi různé, pak podle aplikace.
  - Volantem budu točit vlevo nebo vpravo, ne doprostřed...

Míchání, ADAPT & exponenciální třída

$$p_j(x_t|\Theta) = \exp\left(\langle \eta(\Theta), T(x_t)\rangle - \Psi\left(\eta(\Theta)\right) + k_j(x_t)\right)$$

$$\omega_j^i \in [0,1], \qquad \sum_{j\in\mathcal{N}_i} \omega_j^i = 1.$$

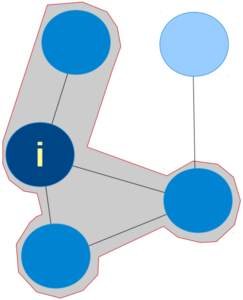
Co bude hezčí na míchání?

Předpoklad rozumných p, ω

$$\tilde{p}_i = \sum_{j\in\mathcal{N}_i} \omega_j^i p_j \qquad \left(\sum \omega_j^i \mathcal{D}(p_j||\tilde{p}_i)\right)$$

nebo

$$\tilde{p}_i \propto \prod_{j\in\mathcal{N}_i} p_j^{\omega_j^i} \qquad \left(\sum \omega_j^i \mathcal{D}(\tilde{p}_i||p_j)\right)$$



Míchání, COMBINE & exponenciální třída

$$\pi_j(\Theta|\xi_{j,t-1},\nu_{j,t-1}) = \exp\left(\langle\eta(\Theta),\xi_{j,t-1}\rangle - \nu_{j,t-1}\Psi\left(\eta(\Theta)\right) + I(\xi_{j,t-1},\nu_{j,t-1})\right)$$

$$\omega_j^i \in [0,1], \qquad \sum_{j\in\mathcal{N}_i} \omega_j^i = 1.$$

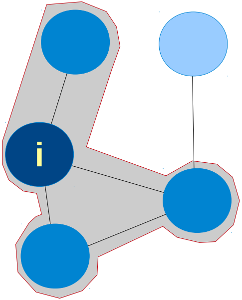
Co bude hezčí na míchání?

Předpoklad rozumných π_j, ω

$$\tilde{\pi}_i = \sum_{j\in\mathcal{N}_i} \omega_j^i \pi_j \qquad \left(\sum \omega_j^i \mathcal{D}(\pi_j||\tilde{\pi}_i)\right)$$

nebo

$$\tilde{\pi}_i \propto \prod_{j\in\mathcal{N}_i} \pi_j^{\omega_j^i} \qquad \left(\sum \omega_j^i \mathcal{D}(\tilde{\pi}_i||\pi_j)\right)$$



Míchání & exponenciální třída & geom. průměr

$$p_j(x_t|\Theta) = \exp\left(\langle \eta(\Theta), T(x_t)\rangle - \Psi\left(\eta(\Theta)\right) + k_j(x_t)\right)$$

$$\pi_j(\Theta|\xi_{j,t-1}, \nu_{j,t-1}) = \exp\left(\langle \eta(\Theta), \xi_{j,t-1}\rangle - \nu_{j,t-1}\Psi\left(\eta(\Theta)\right) + l(\xi_{j,t-1}, \nu_{j,t-1})\right)$$

$$\omega_j^i \in [0,1], \qquad \sum_{j\in\mathcal{N}_i} \omega_j^i = 1.$$

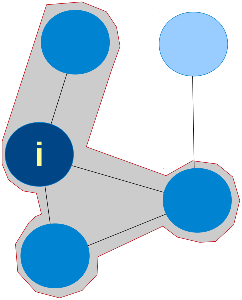
- ADAPT

$$\tilde{T}(x_{i,t}) = \sum_{j\in\mathcal{N}_i} \omega_i^j T(x_j)$$

- COMBINE

$$\tilde{\xi}_{i,t-1} = \sum_{j\in\mathcal{N}_i} \omega_i^j \xi_{j,t-1}$$

$$\tilde{\nu}_{i,t-1} = \sum_{j\in\mathcal{N}_i} \omega_i^j \nu_{j,t-1}$$



Někdy ale konjugovaná apriorna nemáme...

Typický problém: zobecněné lineární modely (GLM)

$$\mathbb{E}[y_t|x_t,\Theta] = g^{-1}(x_t^\intercal \theta)$$

lineární (regresní) model

- g je identita
- ...konjug. apriorno existuje

logistický (regresní) model

- $g(p_t) = \text{logit}(p_t) = \log \frac{p_t}{1-p_t}$
- ...rozumné konj. apriorno neex.

Poměrně často lze použít rozumné nekonj. apriorno ($\mathcal{N}$) a aposteriorno aproximovat např. Laplaceovsky ($\to \mathcal{N}$).

Míchání, COMBINE & Laplaceovské aposteriorno

$$p_j^L(\Theta|\cdot) = \exp\left(\langle \eta(\Theta_j), T(\cdot)\rangle \ldots\right) = \exp\left(\langle \eta_j), T(\cdot)\rangle \ldots\right)$$

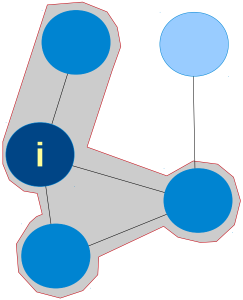
$$\omega_j^i \in [0,1], \qquad \sum_{j\in\mathcal{N}_i} \omega_j^i = 1.$$

Vezmeme-li geom. průměr

$$q \propto \prod_{j\in\mathcal{N}_i} f_j^{\omega_j^i} \qquad \left(\sum \omega_j^i \mathcal{D}(q||f_j)\right)$$

dostaneme snadno

$$\tilde{\eta}_i = \sum_{j\in\mathcal{N}_i} \omega_j^i \eta_j$$



Vlastnosti odhadů

Jde o shrinkage odhad, ležící někde mezi dvěma extrémy

Buď $\omega_\alpha^i = 1$ pro odhad $\pi_\alpha(\Theta|\cdot)$, potom je $\tilde{p}_i = \pi_\alpha(\Theta|\cdot)$.

α je "nejhorší" ve smyslu odhadu

- $\tilde{p}_i$ bude nejhorší odhad

α je "nejlepší" ve smyslu odhadu

- $\tilde{p}_i$ bude nejlepší odhad

Vše ostatní je mezi. Vhodná volba $\omega_j^i \ \forall j \in \mathcal{N}_i$ je tedy kritická.

…ovšem za rozumných podmínek to s tou kritičností není tak žhavé.

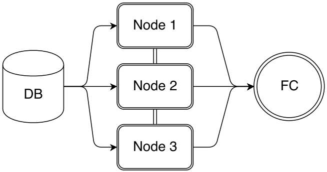
Aplikace: velká dynamická data

Dynamické modelování dynamických velkých dat

- např. v průmyslu (senzorové sítě aj.)
- některé úlohy jsou výpočetně náročné, proto je vhodné je rozdělit
- některé sítě jsou geograficky rozsáhlé nebo mají malou přenosovou rychlost

---

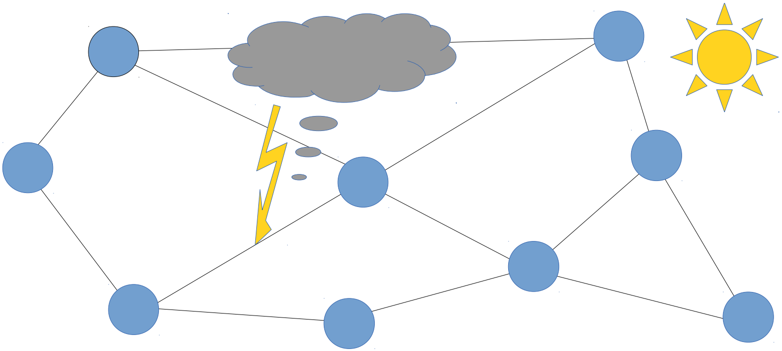
- DB (centralizovaná, decentralizovaná) je FIFO
- Node 1–3 jsou vrcholy grafu
  - berou data na modelování z DB
  - mohou je brát různou rychlostí
  - mohou spolupracovat
- FC je fúzní centrum
  - přistupuje ke všem vrcholům
  - může a nemusí být

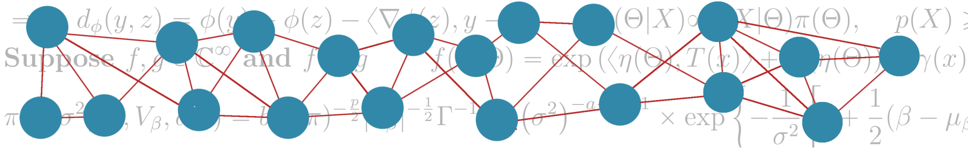


Aplikace: dynamická analýza přežití

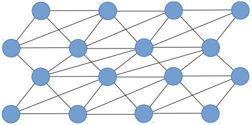
Dynamické modelování vlastností zařízení/sítí

- např. v bezdrátových datových sítích – značná závislost na velikosti paketů
- v různých částech sítě se mohou podmínky lišit (povětrnostní podmínky)
- hledáme optimální velikost paketu pro globální nastavení





A to je vše :-)



http://diffest.utia.cas.cz

GAČR GP14-06678P