

Doplnenie

Supra-bayesovská kombinace pravděpodobnostních distribucí

Vladimíra Sečkárová

- 1 Z_1, \dots, Z_s - vybrané zdroje
- 2 $\mathbf{X} = (X_1, \dots)$ - doména = náhodný vektor s realizáciami $\mathbf{x}_1, \dots, \mathbf{x}_n, n < \infty$
- 3 $(g_j(\mathbf{x}_1), \dots, g_j(\mathbf{x}_n))$ - informácia dodaná j -tým zdrojom vo forme pravdepodobností jednotlivých realizácií, $j = 1, \dots, s$
- 4 optimálny merger:

$$O_{\hat{h}} = \text{Argmin}_{\hat{h}} E_{\pi(h|D)}[K(h, \hat{h})|D] =$$

$$\text{Argmin}_{\hat{h}} K(E_{\pi(h|D)}(h|D), h) = E_{\pi(h|D)}(h|D)$$

- 5 voľba $\pi(h|D)$: pomocou princípu maximálnej entropie s obmedzeniami (podmienkami) na vzdialenosť dodaných pravdepodobností od objektívnych pravdepodobností (t.j. riešim optimalizačný problém vzhľadom k daným podmienkam)
- 6 ${}^O\pi(h|D)$ – riešenie optimalizačného problému:

$$\text{Argmax}_{\pi(h|D) \in M} - \int_M \pi(h|D) \log \pi(h|D) dh,$$

kde

$$M = \{ \pi(h|D) : E_{\pi(h|D)}[K(g_j, h)|D] - \beta_j(D) \leq 0, j = 1, \dots, s,$$

$$\int_H \pi(h|D) dh - 1 = 0, \pi(h|D) > 0 \quad \forall h \in H$$

- 7 minimalizácia odpovedajúceho Lagrangiánu $\rightarrow {}^O\pi(h|D)$ (pre tento vektor pravdepodobností je minimálna Kullbackova Leiblerova divergencia)

- 8 vhodnou voľbou normalizačnej konštanty získame ${}^0\pi(h|D)$ ako hustotu Dirichletovho rozdelenia
- 9 z vlastností Dirichletovho rozdelenia (podmienená stredná hodnota premenných $h(\mathbf{x}_i)$, $i = 1, \dots, n$):

$${}^0\hat{h}(\mathbf{x}_i) = \frac{1}{n + \sum_{j=1}^s \lambda_j(D)} + \sum_{j=1}^s \frac{\lambda_j(D)}{n + \sum_{j=1}^s \lambda_j(D)} g_j(\mathbf{x}_i) \quad (1)$$

- 10 nepravdepodobnostná informácia je dodaná - prevedenie na pravdepodobnostnú:
 – (podmienené) momenty:

$$E_{g_j(\mathbf{f}|\mathbf{p})}(\phi(\mathbf{F}_j, \mathbf{P}_j)|\mathbf{P}_j) = \psi(\mathbf{P}_j)$$

utvorím množinu vyhovujúcich vektorov podmienených pravdepodobností (splňajúcich, že sa podmienené momenty vzhľadom k týmto hustotám rovnajú dodaným údajom), z nich vyberieme vektor podmienených pravdepodobností s najväčšou entropiou

- 11 rozšírenie na zjednotenie domén (v prípade dodania podmienených momentov, t.j. po prevedení: vektorov podmienených pravdepodobností na doméne konkrétneho zdroja):

$$e_{g_j(\mathbf{x}_i)} = {}^O \hat{h}(\mathbf{u}_i|\mathbf{f}_i, \mathbf{p}_i) g_j(\mathbf{f}_i|\mathbf{p}_i) {}^O \hat{h}(\mathbf{p}_i)$$

- 12 ^e g_j dosadiť do vzorca pre optimálny zmiešavač
- 13 spätne premietnuť získaný optimálny odhad (optimálna zmiešavač) na doménu 1. zdroja (i na domény susedov)

Príklad - výsledok

opti1 - voľba β	1000/1085	1000/1196	1000/11364	1000/713
opti2 - voľba β	1/1085	1/1196	1/11364	1/713
opti1 - $\lambda(D)$	0.23	0.25	0.38	0.14
opti2 - $\lambda(D)$	0.25	0.26	0.24	0.24