Univerzita Karlova v Praze

Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE

David Marek

Implementace aproximativních bayesovských metod pro odhad stavu v dialogových systémech

Ústav formální a aplikované lingvistiky

Vedoucí diplomové práce: Ing. Mgr. Filip Jurčíček, Ph.D.

Studijní program: program

Studijní obor: obor

Praha 2013

Poděkování.

Prohlašuji, že jsem tuto diplomovou práci vypracoval(a) samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V .... dne .... Podpis autora

Název práce: Implementace aproximativních bayesovských metod pro odhad stavu v dialogových systémech

Autor: David Marek

Katedra: Ústav formální a aplikované lingvistiky

Vedoucí diplomové práce: Ing. Mgr. Filip Jurčíček, Ph.D., Ústav formální a aplikované lingvistiky

Abstrakt:

Klíčová slova:

Title:

Author: David Marek

Department: Institute of Formal and Applied Linguistics

Supervisor: Ing. Mgr. Filip Jurčíček, Ph.D., Institute of Formal and Applied Linguistics

Abstract:

Keywords:

Obsah

Úvod

Dialog je přirozený způsob dorozumívání a sdělování informací mezi lidmi. Počítač, který by dokázal vést dialog s uživatelem, byl vždy snem nejen příznivců vědecko-fantastické literatury. Už pro první počítače vnikaly programy, které se snažily využívat přirozenou řeč pro interakci s uživatelem. Jedním z takových programů byl například Eliza, program, který předstíral, že jej zajímá, co mu uživatel říká. Fungoval na principu rozpoznání textu pomocí gramatiky a následné transformace textu do promluv dle pravidel. Avšak gramatiky a pravidlové systémy se ukázaly nedostačné pro praktické aplikace a tak se vývoj přesunul do statistických metod. S využitím statistickým metod a metod strojového učení bylo možné začít s porozumíváním mluveného slova. Přijetí bylo zpočátku chladné a veřejnost byla

1. Název první kapitoly

1.1 Název první podkapitoly v první kapitole

1.2 Název druhé podkapitoly v první kapitole

2. Název druhé kapitoly

2.1 Název první podkapitoly v druhé kapitole

2.2 Název druhé podkapitoly v druhé kapitole

3. Učení parametrů

3.1 Grafický model

Máme vybraný faktor f, tento faktor je spojený s několika proměnnými $\boldsymbol{x} = (x_0, x_1, \ldots, x_{N_x})$ a množinami parametrů $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{N_\theta})$. Tento faktor reprezentuje podmíněnou pravděpodobnost:

$$f(\boldsymbol{x}, \Theta) = p(x_0 | x_1, \ldots, x_{N_x}; \Theta)$$

Rodičovské proměnné $x_1, \ldots, x_{N_x}$ označujeme jako $\boldsymbol{x'}$. Vektor $\boldsymbol{x'}$ určuje, která množina parametrů bude použita. Protože množiny parametrů jsou číslovány $1, \ldots, N_\theta$ a rodičovské proměnné $1, \ldots, N_x$, musí být pro vybrání správné množiny parametrů použito mapování $\rho(\boldsymbol{x'})$. Faktor pak může být zapsán zkráceně:

$$f(\boldsymbol{x}, \Theta) = p(x_0 | x_1, \ldots, x_{N_x}; \Theta) = \theta_{\rho(x'), x_0}$$

3.2 Výpočet marginálních pravděpodobností

Pro výpočet sdružené pravděpodobnosti používáme plně faktorizovanou distribuci. Pro každou proměnnou, anebo množinu parametrů je její marginální pravděpodobnost rovna součinu zpráv přicházejících z faktorů, které jsou s danou proměnnou nebo množinu parametrů propojeny. Pro daný faktor je cavity distribuce $q^{\setminus}(x_i)$, popř. $q^{\setminus}(\boldsymbol{\theta}_i)$ rovna součinu všech ostatních faktorů. Aproximovaná marginální pravděpodobnost proměnné je pak součinem cavity distribuce a zprávy z faktoru:

$$q(x_i) = q^{\setminus}(x_i) m_{f \to x_i}(x_i)$$

$$q(\boldsymbol{\theta}_i) = q^{\setminus}(\boldsymbol{\theta}_i) m_{f \to \theta}(\boldsymbol{\theta}_i)$$

3.2.1 Marginální pravděpodobnost proměnných

Pokud chceme aktualizovat hodnotu naší aproximace marginální pravděpodobnosti, tak je třeba minimalizovat její vzdálenost od skutečné marginální pravděpodobnosti:

$$\begin{aligned}
p^*(\tilde{x}_j) &= \sum_{\boldsymbol{x}: x_j = \tilde{x}_j} \int_{\boldsymbol{\Theta}} \prod_i q^{\setminus}(x_i) \prod_l q^{\setminus}(\boldsymbol{\theta}_l) f(\boldsymbol{x}; \boldsymbol{\Theta}) && (3.1) \\
&= \sum_{\boldsymbol{x}: x_j = \tilde{x}_j} \prod_i q^{\setminus}(x_i) \int_{\boldsymbol{\theta}_{\rho(\boldsymbol{x'})}} q^{\setminus}(\boldsymbol{\theta}_{\rho(\boldsymbol{x'})}) \theta_{\rho(\boldsymbol{x'}), x_0} && (3.2) \\
&= \sum_{\boldsymbol{x}: x_j = \tilde{x}_j} \prod_i q^{\setminus}(x_i) \mathbb{E}_{q^{\setminus}}(\theta_{\rho(\boldsymbol{x'}), x_0}) && (3.3)
\end{aligned}$$

Rovnost (3.1) vychází z definice výpočtu marginální pravděpodobnosti ze sdružené pravděpodobnosti. V (3.2) byla použita definice faktoru, z integrálu byly vytaženy členy, které neobsahují Θ a nakonec bylo využito toho, že pro množiny parametrů, které nejsou spojeny s faktorem f, je jejich jejich cavity distribuce rovná marginální distribuci a tedy $\int_{\theta_i} q(\theta_i) = 1$. V (3.3) byla použita definice očekávané hodnoty.

Marginální pravděpodobnost proměnné x_i tedy je

$$p^*(\tilde{x}) = \sum_{\boldsymbol{x}:x_j=\tilde{x}_j} \prod_i q^{\setminus}(x_i)\mathbb{E}_{q^{\setminus}}(\theta_{\rho(\boldsymbol{x'}),x_0}) \tag{3.4}$$

Tady docházíme k výsledku, který je velmi podobný výpočtu marginální pravděpodobnosti v Loopy Belief Propagation algoritmu, střední hodnota $\mathbb{E}_{q^{\setminus}}(\theta_{\rho(\boldsymbol{x'}),x_0})$ zde reprezentuje zprávu z vrcholu $\boldsymbol{\theta}_{\rho_x}$.

Zprávu z faktoru f do vrcholu x_j pak získáme vydělením zprávy z x_j z marginální pravděpodobnosti.

$$m_{f\to x_j}(x_j) = \sum_{\boldsymbol{x}:x_j=\tilde{x}_j} \prod_{i\neq j} q^{\setminus}(x_i)\mathbb{E}_{q^{\setminus}}(\theta_{\rho(\boldsymbol{x'}),x_0}) \tag{3.5}$$

3.2.2 Marginální pravděpodobnost parametrů

Pro množiny parametrů se jejich marginální pravděpodobnost spočítá podobně jako pro proměnné.

$$\begin{aligned}
p^*(\tilde{\boldsymbol{\theta}}_j) &= \sum_{\boldsymbol{x}} \int_{\boldsymbol{\Theta}:\boldsymbol{\theta}_j=\tilde{\boldsymbol{\theta}}_j} \prod_i q^{\setminus}(x_i) \prod_l q^{\setminus}(\boldsymbol{\theta}_l) f(\boldsymbol{x};\boldsymbol{\Theta}) && (3.6)\\
&= \sum_{l\neq j} \sum_{\boldsymbol{x}:\rho(\boldsymbol{x'})=l} \prod_i q^{\setminus}(x_i) \int_{\boldsymbol{\Theta}:\boldsymbol{\theta}_j=\tilde{\boldsymbol{\theta}}_j} \prod_k q^{\setminus}(\boldsymbol{\theta}_k)\theta_{l,x_0} + && (3.7)\\
&+ \sum_{\boldsymbol{x}:\rho(\boldsymbol{x'})=j} \prod_i q^{\setminus}(x_i) \int_{\boldsymbol{\Theta}:\boldsymbol{\theta}_j=\tilde{\boldsymbol{\theta}}_j} \prod_k q^{\setminus}(\boldsymbol{\theta}_k)\tilde{\theta}_{j,x_0} \\
&= \left[\sum_{l\neq j} \sum_{\boldsymbol{x}:\rho(\boldsymbol{x'})=l} \prod_i q^{\setminus}(x_i)\mathbb{E}_{q^{\setminus}(\boldsymbol{\theta}_l)}(\theta_{l,x_0})\right] q^{\setminus}(\tilde{\boldsymbol{\theta}}_j) + && (3.8)\\
&+ \sum_{\boldsymbol{x}:\rho(\boldsymbol{x'})=j} \prod_i q^{\setminus}(x_i)\tilde{\theta}_{j,x_0} q^{\setminus}(\tilde{\boldsymbol{\theta}}_j) \\
&= w_0 q^{\setminus}(\tilde{\boldsymbol{\theta}}_j) + \sum_k w_k \tilde{\theta}_{j,k} q^{\setminus}(\tilde{\boldsymbol{\theta}}_j), && (3.9)
\end{aligned}$$

kde

$$\begin{aligned}
w_0 &= \sum_{l\neq j} \sum_{\boldsymbol{x}:\rho(\boldsymbol{x'})=l} \prod_i q^{\setminus}(x_i)\mathbb{E}_{q^{\setminus}(\boldsymbol{\theta}_l)}(\theta_{l,x_0}) && (3.10)\\
w_k &= \sum_{\boldsymbol{x}:\rho(\boldsymbol{x'})=j,x_0=k} \prod_i q^{\setminus}(x_i) && (3.11)
\end{aligned}$$

Opět vycházíme z výpočtu marginální pravděpodobnosti ze sdružené pravděpodobnosti. V rovnici (3.7) jsme rozdělili sumu přes $\boldsymbol{x}$ na ty, pro které se ve faktoru použije množina parametrů $\tilde{\boldsymbol{\theta}}_j$ a na ty ostatní. Také jsme z integrálu vytknuli součin cavity distribucí pro proměnné. V dalším kroku (3.8) jsme opět použili toho, že integrál přes $\boldsymbol{\Theta}$ je ve skutečnosti několik integrálů přes jednotlivé množiny parametrů. A tedy je můžeme vložit mezi jednotlivé členy produktu cavity distribucí pro množiny parametrů. Ve výsledku získáme $q^{\setminus}(\tilde{\boldsymbol{\theta}}_j) \int_{\boldsymbol{\theta}_l} q^{\setminus}(\boldsymbol{\theta}_l)\theta_{l,x_0}$ a pak zbylé členy, které zmizí.

Docházíme k vyjádření skutečné marginální pravděpodobnosti, ve které není třeba integrovat přes všechny množiny parametrů, ale stačí jen očekávaná hodnota těchto parametrů.

3.3 Aproximace marginálních pravděpodobností

Stále tu ovšem zůstává problém, že spočítat aproximující distribuci $q(\boldsymbol{\theta}_j)$ může být příliš složité, protože skutečná marginální distribuce je směs několika distribucí a ta nemusí být v obecném případě vyjádřitelná. Je tedy třeba model dále aproximovat. Pro zjednodušení výpočtu jsou zprávy z faktoru do množiny parametrů, $m_{f\to\boldsymbol{\theta}_i}(\boldsymbol{\theta}_i)$, ve tvaru Dirichletovského rozdělení s parametry $\hat{\boldsymbol{\alpha}}_i$:

$$m_{f\to\boldsymbol{\theta}_i}(\boldsymbol{\theta}_i) = Dir(\boldsymbol{\theta}_i; \hat{\boldsymbol{\alpha}}_i) = \frac{\Gamma(\sum_j \hat{\alpha}_{i,j})}{\prod_j \Gamma(\hat{\alpha}_{i,j})} \prod_j \theta_{i,j}^{\hat{\alpha}_{i,j}-1} \tag{3.12}$$

kde Γ je Gamma funkce (zobecnění faktoriálu):

$$\Gamma(z) = \int_0^\infty t^{z-1} \exp(-t)\mathrm{d}t \tag{3.13}$$

Dirichletovské rozdělení bylo zvoleno, protože má důležité vlastnosti pro součin, které budou využity dále pro výpočet cavity distribuce a celkové aproximace. Pokud označíme aproximované faktory indexem β a každý bude mít vlastní parametry $\hat{\alpha}_{\beta,i}$, tak výsledná aproximace bude tvaru:

$$\begin{aligned}
q(\boldsymbol{\theta}_i) &\propto \prod_\beta m_{f_\beta\to\boldsymbol{\theta}_i}(\boldsymbol{\theta}_i) && (3.14)\\
&\propto \prod_\beta \prod_j \theta_{i,j}^{\hat{\alpha}_{\beta,i,j}-1} && (3.15)\\
&\propto Dir(\boldsymbol{\theta}_i; \sum_\beta \hat{\boldsymbol{\alpha}}_{\beta,i} - (|\beta|-1)\mathbf{1}) && (3.16)\\
&= Dir(\boldsymbol{\theta}_i; \boldsymbol{\alpha}_i) && (3.17)
\end{aligned}$$

kde $\boldsymbol{\alpha}_i = \sum_\beta \hat{\boldsymbol{\alpha}}_{\beta,i} - (|\beta|-1)\mathbf{1}$).

Při aktualizaci faktoru $\tilde{\beta}$ tedy cavity distribuce bude:

$$\begin{aligned}
q^{\setminus\tilde{\beta}}(\boldsymbol{\theta_i}) &\propto \prod_{\beta\neq\tilde{\beta}} m_{f_\beta\to\boldsymbol{\theta}_i}(\boldsymbol{\theta}_i) && (3.18)\\
&\propto Dir(\boldsymbol{\theta}_i; \boldsymbol{\alpha}_i - \hat{\boldsymbol{\alpha}}_{\beta,i} + \mathbf{1}) && (3.19)
\end{aligned}$$

Naším cílem je nalézt parametry $\boldsymbol{\alpha}^*$ aproximované marginální pravděpodobnosti (3.17), které minimalizují vzdálenost od skutečné marginální pravděpodobnosti (3.9). Pro měření vzdálenosti mezi dvěma pravděpodobnostními rozloženími se používá Kullback-Leiblerova divergence:

$$KL(p\|q) = \int_{-\infty}^{\infty} p(x) \log\left(\frac{p(x)}{q(x)}\right) \mathrm{d}x \tag{3.20}$$

Pro nalezení minima použijeme algoritmus Expectation Propagation a budeme tedy minimalizovat $KL(p^*\|q)$.

Pokud se podíváme na skutečnou marginální pravděpodobnost $p^*(\boldsymbol{\theta}_i)$, zjistíme, že můžeme některé její členy upravit. Využijeme také vlastnosti gamma funkce $\Gamma(x) = (x-1)\Gamma(x-1)$.

$$
\begin{aligned}
w_j\theta_j Dir(\boldsymbol{\theta};\boldsymbol{\alpha}) &\propto w_j\theta_j \frac{\Gamma(\sum_i \alpha_i)}{\prod_i \Gamma(\alpha_i)} \prod_i \theta_i^{\alpha_i - 1} && (3.21)\\
&\propto w_j \frac{\Gamma(\sum_i \alpha_i)}{\prod_i \Gamma(\alpha_i)} \theta_j^{\alpha_j} \prod_{i\neq j} \theta_i^{\alpha_i - 1} && (3.22)\\
&\propto w_j \frac{\Gamma(\sum_i \alpha_i)}{\prod_i \Gamma(\alpha_i)} \frac{\Gamma(\alpha_j + 1)\prod_{i\neq j}\Gamma(\alpha_i)}{\Gamma(1+\sum_i \alpha_i)} Dir(\boldsymbol{\theta};\boldsymbol{\alpha}+\boldsymbol{\delta}_j) && (3.23)\\
&\propto w_j \frac{\Gamma(\sum_i \alpha_i)}{\prod_i \Gamma(\alpha_i)} \frac{\alpha_j\Gamma(\alpha_j)\prod_{i\neq j}\Gamma(\alpha_i)}{(\sum_i \alpha_i)\Gamma(\sum_i \alpha_i)} Dir(\boldsymbol{\theta};\boldsymbol{\alpha}+\boldsymbol{\delta}_j) && (3.24)\\
&\propto w_j \frac{\alpha_j}{\sum_i \alpha_i} Dir(\boldsymbol{\theta};\boldsymbol{\alpha}+\boldsymbol{\delta}_j) && (3.25)\\
& && (3.26)
\end{aligned}
$$

Díky této úpravě lze p^* vyjádřit jako směs Dirichletovských rozdělení.

$$
p^*(\boldsymbol{\theta}) = w_0^* Dir(\boldsymbol{\theta};\boldsymbol{\alpha}) + \sum_j w_j^* Dir(\boldsymbol{\theta};\boldsymbol{\alpha}+\boldsymbol{\delta}_j) \tag{3.27}
$$

kde

$$
\begin{aligned}
w_0^* &\propto w_0 && (3.28)\\
w_j^* &\propto w_j \frac{\alpha_j}{\sum_i \alpha_i} && (3.29)\\
\sum_{i=0}^{k} w_i^* &= 1 && (3.30)
\end{aligned}
$$

Pro minimalizaci KL divergence mezi dvěma rozděleními z exponenciální rozdělení stačí, pokud se budou rovnat jejich postačující statistiky. Dokážeme jednoduše spočítat první dva momenty Dirichletovského rozdělení a tedy použijeme aproximaci a budeme počítat pouze s nimi a zbylé momenty zanedbáme. Je tedy třeba nalézt střední hodnotu a rozptyl proměnných z $p^*(\boldsymbol{\theta})$.

$$
\begin{aligned}
\mathbb{E}_{p^*}[\boldsymbol{\theta}] &= \int \boldsymbol{\theta} p^*(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta} && (3.31)\\
&= \int \boldsymbol{\theta}(w_0^* Dir(\boldsymbol{\theta};\boldsymbol{\alpha}) + \sum_j w_j^* Dir(\boldsymbol{\theta};\boldsymbol{\alpha}+\boldsymbol{\delta}_j))\, \mathrm{d}\boldsymbol{\theta} && (3.32)\\
&= w_0^* \int \boldsymbol{\theta} Dir(\boldsymbol{\theta};\boldsymbol{\alpha})\, d\boldsymbol{\theta} + \sum_j w_j^* \int \boldsymbol{\theta} Dir(\boldsymbol{\theta};\boldsymbol{\alpha}+\boldsymbol{\delta}_j)\, \mathrm{d}\boldsymbol{\theta} && (3.33)\\
&= w_0^* \mathbb{E}_{Dir(\boldsymbol{\alpha})}[\boldsymbol{\theta}] + \sum_j w_j^* \mathbb{E}_{Dir(\boldsymbol{\alpha}+\boldsymbol{\delta}_j)}[\boldsymbol{\theta}] && (3.34)
\end{aligned}
$$

Střední hodnotu proměnných $\boldsymbol{\theta}$ podle rozdělení p^* lze tedy spočítat jako vážený součet středních hodnot $\boldsymbol{\theta}$ podle jednotlivých Dirichletovských distribucí, z kterých se p^* skládá. Střední hodnota proměnné X_i podle Dirichletovského rozdělení je

První moment tedy máme spočítaný, pro výpočet rozptylu můžeme využít přímo definici:

$$Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \tag{3.35}$$

Chybí nám tedy ještě výpočet střední hodnoty druhé mocniny proměnné $\boldsymbol{\theta}$ podle p^*. Můžeme ji vyjádřit z definice střední hodnoty.

$$\begin{aligned}
\mathbb{E}_{p^*}[\boldsymbol{\theta}^2] &= \int \boldsymbol{\theta}^2 p^*(\boldsymbol{\theta})\ d\boldsymbol{\theta} && (3.36)\\
&= w_0^* \int \boldsymbol{\theta}^2 Dir(\boldsymbol{\theta};\boldsymbol{\alpha})\ d\boldsymbol{\theta} + \sum_j w_j^* \int \boldsymbol{\theta}^2 Dir(\boldsymbol{\theta};\boldsymbol{\alpha}+\boldsymbol{\delta}_j)\ d\boldsymbol{\theta} && (3.37)\\
&= w_0^* \mathbb{E}_{Dir(\boldsymbol{\alpha})}[\boldsymbol{\theta}^2] + \sum_j w_j^* \mathbb{E}_{Dir(\boldsymbol{\alpha}+\boldsymbol{\delta}_j)}[\boldsymbol{\theta}^2] && (3.38)
\end{aligned}$$

Opět získáváme vážený součet středních hodnot podle Dirichletovských rozdělení. Střední hodnotu druhé mocniny proměnné podle Dirichletovského rozdělení lze opět jednoduše odvodit z definice.

$$\begin{aligned}
\mathbb{E}_{Dir(\boldsymbol{\alpha})}[x_i^2] &= \int x_i^2 Dir(\boldsymbol{x};\boldsymbol{\alpha}) d\boldsymbol{x} && (3.39)\\
&= \int x_i^2 \frac{\Gamma(\alpha_0)}{\prod_{j=1}^N \Gamma(\alpha_j)} \prod_{j=1}^N x_j^{\alpha_j - 1} d\boldsymbol{x} && (3.40)
\end{aligned}$$

Nyní jsme ve stejné situaci jako v (3.21). Budeme postupovat stejně, vyjádříme nové Dirichletovské rozdělení.

$$\begin{aligned}
\mathbb{E}_{Dir(\boldsymbol{\alpha})}[x_i^2] &= \int \frac{\Gamma(\alpha_0+2)\alpha_i(\alpha_i+1)}{\alpha_0(\alpha_0+1)\Gamma(\alpha_i+2)\prod_{j\neq i}\Gamma(\alpha_j)} x_i^{\alpha_i+1} \prod_{j\neq i} x_j^{\alpha_j-1} \mathrm{d}\boldsymbol{x} && (3.41)\\
&= \frac{\alpha_i(\alpha_i+1)}{\alpha_0(\alpha_0+1)} \int \frac{\Gamma(\beta_0)}{\prod_i \Gamma(\beta_i)} \prod_i x_i^{\beta_i-1} \mathrm{d}\boldsymbol{x} && (3.42)\\
&= \frac{\alpha_i(\alpha_i+1)}{\alpha_0(\alpha_0+1)} \int Dir(\boldsymbol{x};\boldsymbol{\beta}) \mathrm{d}\boldsymbol{x} && (3.43)\\
&= \frac{\alpha_i(\alpha_i+1)}{\alpha_0(\alpha_0+1)} && (3.44)
\end{aligned}$$

Vyjádřili jsme $\Gamma(\alpha_0)$ a $\Gamma(\alpha_i)$ s pomocí $\Gamma(\alpha_0+2)$ a $\Gamma(\alpha_i+2)$

$$\Gamma(\alpha_0) = \frac{\Gamma(\alpha_0+2)}{\alpha_0(\alpha_0+1)} \tag{3.45}$$

$$\Gamma(\alpha_i) = \frac{\Gamma(\alpha_i+2)}{\alpha_i(\alpha_i+1)} \tag{3.46}$$

Následně jsme vytvořili nové parametry $\boldsymbol{\beta}$:

$$\beta_i = \alpha_i + 2 \tag{3.47}$$
$$\beta_{j \neq i} = \alpha_j \tag{3.48}$$
$$\beta_0 = \sum_i \beta_i \tag{3.49}$$

Nyní tedy dokážeme spočítat $\mathbb{E}_{p^*}[\boldsymbol{\theta}]$ a $\mathbb{E}_{p^*}[\boldsymbol{\theta}^2]$. Parametry aproximovaného rozdělení nalezneme následovně

$$\frac{\mathbb{E}[X_1] - \mathbb{E}[X_1^2]}{\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2} = \frac{\frac{\alpha_1}{\alpha_0} - \frac{\alpha_1(\alpha_1+1)}{\alpha_0(\alpha_0+1)}}{\frac{\alpha_1(\alpha_1+1)}{\alpha_0(\alpha_0+1)} - \frac{\alpha_1^2}{\alpha_0^2}} \tag{3.50}$$
$$= \frac{\frac{\alpha_1(\alpha_0+1)-\alpha_1(\alpha_1+1)}{\alpha_0(\alpha_0+1)}}{\frac{\alpha_0\alpha_1(\alpha_1+1)-\alpha_1^2(\alpha_0+1)}{\alpha_0^2(\alpha_0+1)}} \tag{3.51}$$
$$= \frac{\alpha_0\alpha_1(\alpha_0 - \alpha_1)}{\alpha_1(\alpha_0\alpha_1 + \alpha_0 - \alpha_0\alpha_1 - \alpha_1)} \tag{3.52}$$
$$= \alpha_0 \tag{3.53}$$
$$\alpha_i = \mathbb{E}[X_i]\alpha_0 \tag{3.54}$$

Z rovnice (3.50) vypočítáme sumu všech parametrů α_0, protože střední hodnota proměnné z Dirichletovského rozdělení je právě $\frac{\alpha_i}{\alpha_0}$, tak jednotlivé parametry získáme z rovnice (3.54).

3.4 Algoritmus

Algoritmus 1 Expectation Propagation pro učení parametrů

Parametry zpráv z faktoru β do množiny parametrů $\boldsymbol{\theta}_i$ označíme $\hat{\boldsymbol{\alpha}}_{\beta,i}$.
Parametry zpráv z množiny parametrů $\boldsymbol{\theta}_i$ do faktoru β označíme $\boldsymbol{\alpha}_i^{\beta}$.
Parametry marginální distribuce množiny parametrů $\boldsymbol{\theta}_i$ označíme $\boldsymbol{\alpha}_i$.
init
  Nastav zprávy mezi faktory a proměnnými na 1.
  Nastav parametry $\hat{\boldsymbol{\alpha}}_{\beta,i}$ na 1.
  Nastav parametry $\boldsymbol{\alpha}_i$ na apriorní hodnotu.
end init
repeat
  Vyber faktor $f_{\tilde{\beta}}$, který se bude aktualizovat.
  Spočítej všechny zpravy z parametrů:
  for každý parametr $\boldsymbol{\theta}_i$ spojený s faktorem $f_{\tilde{\beta}}$ **do**
    Parametry zprávy z $\boldsymbol{\theta}_i$ do $f_{\tilde{\beta}}$: $\boldsymbol{\alpha}_i^{\tilde{\beta}} = \boldsymbol{\alpha}_i - \hat{\boldsymbol{\alpha}}_{\tilde{\beta},i} + 1$.
  end for
  Aktualizuj zprávy z faktoru do proměnných:
  for každou proměnnou X_i, spojenou s faktorem $f_{\tilde{\beta}}$ **do**
    Zpráva z $f_{\tilde{\beta}}$ do X_i podle (3.5):
      $\hat{f}(x_j) = \sum_{\boldsymbol{x}:x_j=\tilde{x}_j} \mathbb{E}_{q^{\setminus}}(\theta_{\rho(\boldsymbol{x'}),x_0}) \prod_{i\neq j} q^{\setminus}(x_i)$
  end for
  Aktualizuj marginální pravděpodobnost parametrů:
  for každý parametr $\boldsymbol{\theta}_i$ spojený s faktorem $f_{\tilde{\beta}}$ **do**
    Spočítej parametry $\boldsymbol{\alpha}_i^*$ pro Dirichletovské rozdělení, které nejlépe aproximuje cílovou marginální distribuci (3.27). Metoda popsána v sekci **??**
    Parametry zprávy z $f_{\tilde{\beta}}$ do $\boldsymbol{\theta}_i$:
      $\hat{\boldsymbol{\alpha}}_{\beta,i} = \boldsymbol{\alpha}_i^* - \boldsymbol{\alpha}_i^{\setminus\tilde{\beta}} + 1$
    Aktualizuj parametry marginální distribuce $q(\boldsymbol{\theta}_i)$
      $\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_i^* = \hat{\boldsymbol{\alpha}}_{\tilde{\beta},i} + \boldsymbol{\alpha}_i^{\setminus\beta}$.
  end for
  for každou proměnnou X_i, spojenou s faktorem $f_{\tilde{\beta}}$ **do**
    Aktualizuj zprávy z proměnných do faktoru:
      $q^{\setminus\beta}(x_i) = \prod_{\beta\neq\tilde{\beta}} \hat{f}_{\beta}(x_l)$
  end for
until konvergence

Závěr

Seznam použité literatury

Seznam tabulek

Seznam použitých zkratek

Přílohy