REPRODUCIBILITY

February 8, 2021

1 Derivation of ELBO

In this section, we derive the ELBO for VRBot as following, the learning objective from 1-st turn to the t-th turn is to maximize the following formula,

$$\begin{aligned}
&\log \prod_{i=1}^{t} p_\theta(R_i|R_{i-1}, U_i, G^{global}) \\
=&\log \prod_{i=1}^{t} \sum_{S_i, A_i} p_{\theta_g}(R_i|S_i, A_i, R_{i-1}, U_i) \cdot p_{\theta_s}(S_i) \cdot p_{\theta_a}(A_i) \\
=&\log \prod_{i=1}^{t-1} \sum_{S_i, A_i} p_{\theta_g}(R_i|S_i, A_i, R_{i-1}, U_i) \cdot p_{\theta_s}(S_i) \cdot p_{\theta_a}(A_i) \\
&\cdot \Big[\sum_{S_t, A_t} p_{\theta_g}(R_t|S_t, A_t, R_{t-1}, U_t) \cdot p_{\theta_s}(S_t) \cdot p_{\theta_a}(A_t) \Big] \\
=&\log \prod_{i=1}^{t-1} \sum_{S_i, A_i} p_\theta(R_i|R_{i-1}, U_i, G^{global}) \cdot p_\theta(S_i, A_i|S_{i-1}, R_{i-1}, U_i, G^{global}, R_i) \\
&\cdot \Big[\sum_{S_t, A_t} p_{\theta_g}(R_t|S_t, A_t, R_{t-1}, U_t) \cdot p_{\theta_s}(S_t) \cdot p_{\theta_a}(A_t) \Big] \text{(Bayes' rule)},
\end{aligned} \tag{1}$$

where $p_\theta(S_i, A_i|S_{i-1}, R_{i-1}, U_i, G^{global}, R_i)$ is the joint posterior distribution over S_i, A_i, we reach the above transition via Bayes' rule. As we leverage distribution $q_{\phi_s}(S_t) \cdot q_{\phi_a}(A_t)$ to approximate the true posterior distribution, we have the following approximate equality,

$$\begin{aligned}
&\log \prod_{i=1}^{t} p_\theta(R_i|R_{i-1}, U_i, G^{global}) \\
\approx&\log \prod_{i=1}^{t-1} \sum_{S_i, A_i} p_\theta(R_i|R_{i-1}, U_i, G^{global}) \cdot q_{\phi_s}(S_i) \cdot q_{\phi_a}(A_i) \\
&\cdot \Big[\sum_{S_t, A_t} p_{\theta_g}(R_t|S_t, A_t, R_{t-1}, U_t) \cdot p_{\theta_s}(S_t) \cdot p_{\theta_a}(A_t) \Big],
\end{aligned} \tag{2}$$

Where $q_{\phi_s}(S_i)$ and $q_{\phi_a}(A_i)$ are the approximate posterior distributions over S_i and A_i, respectively. In the t-th turn, $U_{\leq t}$ and $R_{\leq t-1}$ are given, thus we omit the marginal distribution $p_\theta(R_i|R_{i-1}, U_t, G^{global})$ if $i \leq t-1$. We can derive the following formula:

$$
\begin{aligned}
&\log \prod_{i=1}^{t} p_\theta(R_i|R_{i-1}, U_i, G^{global}) \\
\approx &\log \prod_{i=1}^{t-1} \sum_{S_i, A_i} p_\theta(R_i|R_{i-1}, U_i) \cdot q_{\phi_s}(S_i) \cdot q_{\phi_a}(A_i) \\
&\cdot \Big[\sum_{S_t, A_t} p_{\theta_g}(R_t|S_t, A_t, R_{t-1}, U_t) \cdot p_{\theta_s}(S_t) \cdot p_{\theta_a}(A_t) \Big], \\
= &\log \prod_{i=1}^{t-1} \sum_{S_i, A_i} q_{\phi_s}(S_i) \cdot q_{\phi_a}(A_i) \\
&\cdot \Big[\sum_{S_t, A_t} p_{\theta_g}(R_t|S_t, A_t, R_{t-1}, U_t) \cdot p_{\theta_s}(S_t) \cdot p_{\theta_a}(A_t) \Big] \\
&(U_{\leq t} \text{ and } R_{\leq t-1} \text{ are given}),
\end{aligned}
\tag{3}
$$

As the state follows the homogeneous Markov hypothesis, which means t-th turn state (i.e., S_t) only relies on the previous turn state (i.e., S_{t-1}). Besides, A_t only relies on S_t, then we have,

$$
\begin{aligned}
&\log p_\theta(R_t|R_{t-1}, U_t, G^{global}) \\
= &\log \sum_{S_{t-1}, A_{t-1}} q_{\phi_s}(S_{t-1}) \cdot q_{\phi_a}(A_{t-1}) \\
&\cdot \Big[\sum_{S_t, A_t} p_{\theta_g}(R_t|S_t, A_t, R_{t-1}, U_t) \cdot p_{\theta_s}(S_t) \cdot p_{\theta_a}(A_t) \Big] \\
&(p_{\theta_s}(S_t) \text{only relies on } S_{t-1}) \\
= &\log \sum_{S_{t-1}} q_{\phi_s}(S_{t-1}) \\
&\cdot \Big[\sum_{S_t, A_t} p_{\theta_g}(R_t|S_t, A_t, R_{t-1}, U_t) \cdot p_{\theta_s}(S_t) \cdot p_{\theta_a}(A_t) \Big] \\
&(p_{\theta_a}(A_t) \text{ only relies on } S_t),
\end{aligned}
\tag{4}
$$

At the t-th turn, following [Kingma and Welling, 2013], we derive the ELBO as following:

$$
\begin{aligned}
&\log p_\theta(R_t|R_{t-1}, U_t, G^{global}) \\
\approx &E_{q_{\phi_s}(S_{t-1})} \big[\log E_{p_{\theta_s}(S_t) \cdot p_{\theta_a}(A_t)} [p_{\theta_g}(R_t|S_t, A_t, R_{t-1}, U_t)] \big] \\
\geq &E_{q_{\phi_s}(S_{t-1})} \big[E_{q_{\phi_s}(S_t) \cdot q_{\phi_a}(A_t)} [\log p_{\theta_g}(R_t|R_{t-1}, U_t, S_t, A_t)] \\
&\qquad - KL(q_{\phi_s}(S_t) \cdot q_{\phi_a}(A_t)) || p_{\theta_s}(S_t) \cdot p_{\theta_a}(A_t) \big], \\
= &- \mathcal{L}_{joint}
\end{aligned}
\tag{5}
$$

where $q_{\phi_s}(S_{t-1})$ is the abbreviation of $q_{\phi_s}(S_{t-1}|S_{t-2}, R_{t-2}, U_{t-1}, R_{t-1})$. Besides, we only execute $p_{\theta_s}(S_t)$ and $p_{\theta_a}(A_t)$ to infer state S_t and action A_t, then send it to $p_{\theta_g}(R_t|\cdot)$ to generate response R_t.

1.1 Derivation of 2-stage collapsed inference

In the training stage, we optimize the state tracker and policy network separately instead of optimize all models jointly. We decompose the training process into two stages, i.e., *state optimization stage* and *action optimization stage*. In the state optimization stage, we ignore to estimate the action posterior distribution, and only use the prior policy network in VRBot, then we have,

$$
\begin{aligned}
&\log p_\theta(R_t|R_{t-1}, U_t, G^{global}) \\
\approx& E_{q_{\phi_s}(S_{t-1})}\big[\log E_{p_{\theta_s}(S_t)\cdot p_{\theta_a}(A_t)}[p_{\theta_g}(R_t|S_t, A_t, R_{t-1}, U_t)]\big] \\
=& E_{q_{\phi_s}(S_{t-1})}\Big[E_{p_{\theta_s}(S_t)}\big[\log E_{p_{\theta_a}(A_t)}[p_{\theta_g}(R_t|S_t, A_t, R_{t-1}, U_t)]\big]\Big] \\
\geq& E_{q_{\phi_s}(S_{t-1})}\Big[E_{q_{\phi_s}(S_t)}\big[\log E_{p_{\theta_a}(A_t)}[p_{\theta_g}(R_t|S_t, A_t, R_{t-1}, U_t)]\big] \\
&- D_{KL}(q_{\phi_s}(S_t)||p_{\theta_s}(S_t))\Big] \\
=& E_{q_{\phi_s}(S_{t-1})}\Big[E_{q_{\phi_s}(S_t)}\big[E_{p_{\theta_a}(A_t)}[\log p_{\theta_g}(R_t|S_t, A_t, R_{t-1}, U_t)]\big] \\
&- D_{KL}(q_{\phi_s}(S_t)||p_{\theta_s}(S_t))\Big] \\
=& -\mathcal{L}_s
\end{aligned}
\tag{6}
$$

Likewise, in the action optimization stage, we derive the ELBO for approximating the state posterior distribution. Illustrated as follows:

$$
\begin{aligned}
&\log p_\theta(R_t|R_{t-1}, U_t, G^{global}) \\
\approx& E_{q_{\phi_s}(S_{t-1})}\big[\log E_{p_{\theta_s}(S_t)\cdot p_{\theta_a}(A_t)}[p_{\theta_g}(R_t|S_t, A_t, R_{t-1}, U_t)]\big] \\
=& E_{q_{\phi_s}(S_{t-1})}\Big[E_{p_{\theta_s}(S_t)}\big[E_{p_{\theta_a}(A_t)}[\log p_{\theta_g}(R_t|S_t, A_t, R_{t-1}, U_t)]\big]\Big] \\
\geq& E_{q_{\phi_s}(S_{t-1})}\Big[E_{p_{\theta_s}(S_t)}\big[E_{q_{\phi_a}(A_t)}[\log p_{\theta_g}(R_t|S_t, A_t, R_{t-1}, U_t)] \\
&- D_{KL}(q_{\phi_a}(A_t)||p_{\theta_a}(A_t))\big]\Big] \\
=& -\mathcal{L}_a
\end{aligned}
\tag{7}
$$

2 Generation Demonstration

One case extracted from KaMed to compare the performance of VRBot, SEDST and PostKS, we list out three rounds as shown in Tab. 1.

References

[Kingma and Welling, 2013] Kingma, D. P. and Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Table 1: One case extracted from KaMed in English and Chinese. ✘,✔ and ✔denote the response is incorrect, moderate and excellent, respectively.

One Case Extracted From KaMed in Chinese and English	
	1st turn
U_1 :	是过敏性鼻炎么（ 女，19 岁），每一次跑步完鼻子特痒，流清鼻涕。(Is it allergic rhinitis (female, 19 years old)? My nose is very itchy and runny after every run.)
R_1:	这情况多久了？有没有打喷嚏？鼻塞？(How long has this been? Did you sneeze? Stuffy nose?)
S_1:	过敏性鼻炎 (allergic rhinitis)
SEDST:	过敏性鼻炎有多长时间了？(How long have you had allergic rhinitis?) ✔
A_1:	过敏性鼻炎 打喷嚏 (allergic rhinitis, sneezing)
PostKS:	是过敏性鼻炎，打喷嚏么？(It is allergic rhinitis, sneezing?) ✔
S_1 :	过敏性鼻炎 (allergic rhinitis)
G_n^{local}:	Allergic rhinitis → symptom → Sneezing (0.07) ; Allergic rhinitis → tsymptom → Stuffy nose (0.02)
A_1 :	询问症状 分泌物 打喷嚏 (ask symptom, secretions, sneezing)
VRBot:	有分泌物吗？打喷嚏吗？(Do you have any secretions? Sneezing?) ✔
	2nd turn
U_2 :	嗯,平时也会感觉鼻子会有一点不舒服，经常感觉鼻子干。(Um. I usually feel the nose be a little uncomfortable, often dry nose.)
R_2:	您可以用雷诺考特喷鼻。(You can use Budesonide Nasal Spray.)
S_1:	打喷嚏 过敏性鼻炎 鼻塞 (Sneezing, Allergic rhinitis, Nasal congestion)
SEDST:	您的情况考虑是过敏性鼻炎引起的，建议您去医院检查一下 (Your symptoms are considered to be caused by allergic rhinitis, it is recommended that you go to the hospital for examination) ✔
A_2:	消炎药 (Anti-inflammatory drugs)
PostKS:	建议你先吃点消炎药。(I suggest you take some anti-inflammatory drugs first.) ✘
S_2 :	过敏性鼻炎 打喷嚏 鼻塞 (Sneezing, Allergic rhinitis, Nasal congestion)
G_n^{local}:	Allergic rhinitis → treated_by → Loratadine (0.02); Allergic rhinitis → treated_by → Montelukast (0.02)
A_2 :	开药 扑尔敏 氯雷他定 (prescribe medicine, Chloramphenicol, Loratadine)
VRBot:	您好，您的情况考虑是过敏性鼻炎引起的，您用过什么药物治疗呢，氯雷他定？(Hello, your symptoms are considered to be caused by allergic rhinitis. What drugs have you used? loratadine?) ✔
	3rd turn
U_3 :	我这种情况还有什么建议吗？(Do you have any suggestions for my symptoms?)
R_3:	生理性海水冲洗鼻腔，口服开瑞坦。(Flushed your nasal cavity with physiological seawater, take loratadine tablets.)
S_3:	血管运动性鼻炎 过敏性鼻炎 鼻塞 打喷嚏 雷诺考特 (Vasomotor rhinitis, Allergic rhinitis, Stuffy nose, Sneezing, Budesonide Nasal Spray)
SEDST:	雷诺考特喷鼻。(Spray your nose with Budesonide Nasal Spray.) ✘
A_3:	过敏性鼻炎 (allergic rhinitis)
PostKS:	您的情况是过敏性鼻炎引起的，建议去医院检查。(Your symptom is caused by allergic rhinitis, suggest you go to the hospital to check the nose.) ✘
S_3 :	血管运动性鼻炎 过敏性鼻炎 鼻塞 打喷嚏(Vasomotor rhinitis, Allergic rhinitis, Stuffy nose, Sneezing)
G_n^{local}:	Allergic rhinitis → treated_by → Montelukast (0.09) ; Allergic rhinitis → treated_by → Cetirizine (0.04); Allergic rhinitis → treated_by → Dexamethasone (0.02)
A_3 :	开药 内舒拿 孟鲁司特(prescribe medicine, Nasonex, Montelukast)
VRBot:	内舒拿喷鼻，口服孟鲁司特钠咀嚼片 (Spray your nose with Nasonex , take Montelukast Sodium Chewable Tablets) ✔