Deep Generative Models with Learnable Knowledge Constraints

Zhiting Hu\textsuperscript{1,2}, Zichao Yang\textsuperscript{1}, Ruslan Salakhutdinov\textsuperscript{1}, Xiaodan Liang\textsuperscript{1}, Lianhui Qin, Haoye Dong, Eric P. Xing\textsuperscript{1,2}
Carnegie Mellon University\textsuperscript{1}
Petuum Inc\textsuperscript{2}
Recent advances in deep generative models

• Deep generative models (DGMs)
  • Variational autoencoders (VAEs) [Kingma & Welling, 2013]
  • Generative adversarial networks (GANs) [Goodfellow et al., 2014]
  • Auto-regressive models

• Impressive success in vision/text domains
Recent advances in deep generative models

• Deep generative models (DGMs)
  • Variational autoencoders (VAEs) [Kingma & Welling, 2013]
  • Generative adversarial networks (GANs) [Goodfellow et al., 2014]
  • Auto-regressive models

• Impressive success in vision/text domains
Recent advances in deep generative models

- Deep generative models
- Variational autoencoders (VAEs) [Kingma & Welling, 2013]
- Generative adversarial networks (GANs) [Goodfellow et al., 2014]
- Auto-regressive models

Impressive success in vision/text domains

- GAN-generated face images by NVIDIA [Karras et al., 2018]
- Machine-written text by OpenAI [Radford et al., 2019]

Legolas and Gimli advanced on the orcs, raising their weapons with a harrowing war cry.

The orcs' response was a deafening onslaught of claws, claws, and claws; even Elrond was forced to retreat. "You are in good hands, dwarf," said Gimli, who had been among the first to charge at the orcs; it took only two words before their opponents were reduced to a blood-soaked quagmire, and the dwarf took his first kill of the night. The battle lasted for hours until two of the largest Orcs attempted to overwhelm Aragorn. When they finally stopped, they lay defeated and lifeless for miles and miles.

"I take nothing," said Aragorn. "But I give my word, at my peril and mine, that I will never forget this day of horror. None of us will forget. Ever!"

"I'll never forget it!" cried Gimli, who had been in the thick of the battle but hadn't taken part in it. One of the wounded Orcs he had carried off, he was the only one of the survivors who remained uninjured. "We'll keep the memory of that day of evil, and the war with it, alive as long as we live, my friends!"

"Then we'll keep it alive as long as we live," added Legolas. "And we won't forget the first great battle of the night, even if we may have forgotten the final defeat."

"I agree," Gandalf said, "but we will all remember it as the last battle in Middle-earth, and the first great battle of the new day."

Aragorn drew his sword, and the Battle of Fangorn was won. As they marched out through the thicket the morning mist cleared, and the day turned to dusk.

The Two Rings were returned to Rivendell. Frodo and Sam woke up alone in their room, and Frodo found a note on his pillow. He opened it and read:

May the Power of the Ring be with you always, and may its light never fade. I am not sure if it matters which of the two rings we accept this day but, as you asked me, I have chosen mine. I am sorry to leave you, Frodo, but know that we are very close to the end, and that you are with us forever. May this letter find you safely in Rivendell; and if it does not, then I will accept the ring in your stead. If by any chance you find or give this letter to the enemy, may they learn the strength of the ring and may the Two Rings never be broken!
Difficulty in exploit problem structures and domain knowledge

• Pose Conditional Person Image Generation
  • Given a person image and a target pose, generate an image of the person under the new pose.
Difficulty in exploiting problem structures and domain knowledge

• Pose Conditional Person Image Generation
  • Given a person image and a target pose, generate an image of the person under the new pose.

• Generative models can be trained on supervised data.
Difficulty in exploit problem structures and domain knowledge

• Pose Conditional Person Image Generation
  • Given a person image and a target pose, generate an image of the person under the new pose

• Generative models can be trained on *supervised data*

• But fail to use *structured knowledge*, i.e., human body structure
  • head, main body, arms, …

![Diagram of Pose Conditional Person Image Generation](image)
Existing approaches to adding structured knowledge

• Designing specialized neural architectures to hard-code knowledge
  • E.g., ConvNets conv-pooling architecture: translation-invariance of image classification
  • Usually only applicable to specific knowledge, models, or tasks
Existing approaches to adding structured knowledge

- **Posterior regularization (PR)** [Ganchev et al., 10; Hu et al., 16]
  - Imposes knowledge constraints on posterior distributions of probabilistic models
  - Many DGMs lack probabilistic Bayesian formulation / meaningful latent variables
  - Require *a priori* fixed constraints
    - Heavy engineering burden / sub-optimal without adaptivity to the data and models

Diagram:
- Source image
- Target pose
- True target
- Generated image
- Human part parser
- Constraint

Structured consistency
Existing approaches to adding structured knowledge

• **Posterior regularization (PR)** [Ganchev et al., 10; Hu et al., 16]
  • Imposes knowledge constraints on posterior distributions of probabilistic models
  • Many DGMs lack probabilistic Bayesian formulation / meaningful latent variables
  • Require *a priori* fixed constraints
    • Heavy engineering burden / sub-optimal without adaptivity to the data and models
This work: DGMs with learnable knowledge

• A general means of incorporating arbitrary structured knowledge with any types of deep (generative) models in a principled way
  • Formal connections between PR and reinforcement learning (RL)
  • Extends PR to learn constraints as the extrinsic reward in RL
This work: DGMs with learnable knowledge

- A general means of incorporating **arbitrary structured knowledge** with any types of deep (generative) models in a principled way
  - Formal connections between PR and *reinforcement learning (RL)*
  - Extends PR to learn constraints as the extrinsic reward in RL
This work: DGMs with learnable knowledge

• A general means of incorporating **arbitrary structured knowledge** with **any types of deep (generative) models** in a principled way
  • Formal connections between **PR** and **reinforcement learning (RL)**
  • Extends PR to **learn constraints** as the extrinsic reward in RL
Posterior Regularization for DGMs

• Consider a generative model $x \sim p_\theta(x)$

• Consider a constraint function $f(x) \in \mathbb{R}$
  • higher $f$ value, better $x$ w.r.t. the knowledge

• PR assumes a variational distribution $q$, and the objective
  $$\min_{\theta, q} \mathcal{L}(\theta, q) = \text{KL}(q(x) || p_\theta(x)) - \alpha \mathbb{E}_q[f(x)]$$
  • Solve with an EM-style procedure
    - **E-step:** $q^*(y|x) \propto p_\theta(x) \exp\{\alpha f(x)\}$
    - **M-step:** $\min_{\theta} \text{KL}(q^*(x)||p_\theta(x)) = \min_{\theta} -\mathbb{E}_{q^*}[\log p_\theta(x)] + \text{const.}$
Posterior Regularization for DGMs

• Consider a generative model \( x \sim p_\theta(x) \)
• Consider a constraint function \( f(x) \in \mathbb{R} \)
  • higher \( f \) value, better \( x \) w.r.t. the knowledge
• PR assumes a variational distribution \( q \), and the objective
  \[
  \min_{\theta, q} \mathcal{L}(\theta, q) = \text{KL}(q(x) \| p_\theta(x)) - \alpha \mathbb{E}_q[f(x)]
  \]
• Solve with an EM-style procedure
  \textbf{E-step:} \( q^*(y|x) \propto p_\theta(x)\exp\{\alpha f(x)\} \)
  \textbf{M-step:} \( \min_\theta \text{KL}(q^*(x) \| p_\theta(x)) = \min_\theta -\mathbb{E}_{q^*}[\log p_\theta(x)] + \text{const.} \)
Posterior Regularization for DGMs

• Consider a generative model \( x \sim p_\theta(x) \)
• Consider a constraint function \( f_\phi(x) \in \mathbb{R} \)
  • higher \( f \) value, better \( x \) w.r.t. the knowledge
• PR assumes a variational distribution \( q \), and the objective
  \[
  \min_{\theta, q} \mathcal{L}(\theta, q) = \text{KL}(q(x) || p_\theta(x)) - \alpha \mathbb{E}_q[f_\phi(x)]
  \]
  • Solve with an EM-style procedure
    \[
    \textbf{E-step: } q_\phi(y|x) \propto p_\theta(x) \exp\{ \alpha f_\phi(x) \}
    \]
    \[
    \textbf{M-step: } \min_\theta \text{KL} \left( q_\phi(x) || p_\theta(x) \right) = \min_\theta -\mathbb{E}_{q_\phi} [\log p_\theta(x)] + \text{const.}
    \]
Posterior Regularization for DGMs

• Consider a generative model $\mathbf{x} \sim p_{\theta}(\mathbf{x})$
• Consider a constraint function $f_{\phi}(\mathbf{x}) \in \mathbb{R}$
  • higher $f$ value, better $\mathbf{x}$ w.r.t. the knowledge
• PR assumes a variational distribution $q$, and the objective
  \[ \min_{\theta,q} \mathcal{L}(\theta, q) = \text{KL}(q(\mathbf{x})|| p_{\theta}(\mathbf{x})) - \alpha \mathbb{E}_{q} [f_{\phi}(\mathbf{x})] \]
  • Solve with an EM-style procedure
    \[ \textbf{E-step: } q_{\phi}(\mathbf{y}|\mathbf{x}) \propto p_{\theta}(\mathbf{x})\exp\{\alpha f_{\phi}(\mathbf{x})\} \]
    \[ \textbf{M-step: } \min_{\theta} \text{KL}\left(q_{\phi}(\mathbf{x})||p_{\theta}(\mathbf{x})\right) = \min_{\theta} -\mathbb{E}_{q_{\phi}} \left[ \log p_{\theta}(\mathbf{x}) \right] + \text{const.} \]

Optimize $\phi$ in M-step: \[ \max_{\phi} \mathbb{E}_{\mathbf{x} \sim q}[f_{\phi}(\mathbf{x})] \]

vanilla PR: constraint $f$ is fixed \textit{a priori}

We want to allow learnable components, i.e. $f_{\phi}$

problematic as the generated samples $\mathbf{x} \sim q$ can be of low quality
Entropy-Regularized Policy Optimization (ERPO)

• ERPO:
  • Policy gradient with information theoretic regularizers
  • E.g., Relative Entropy Policy Search [Peters et al., 2010]

• In RL convention: assume state $s$, action $a$, policy $p_\pi(a|s)$, reward $R(s, a) \in \mathbb{R}$; $\mu^\pi$ is the stationary state distribution

• To map to PR: Let $x = (s, a)$, $p_\pi(x) = \mu^\pi(s)p_\pi(a|s)$

• Let $q_\pi(x)$ be policy at iteration $t$ and $p_\pi(x)$ at iteration $t - 1$

• Objective

$$\min_{q_\pi} \mathcal{L}(q_\pi) = \text{KL}(q_\pi(x) || p_\pi(x)) - \alpha \mathbb{E}_q [R(x)]$$
Close resemblance b/w PR and ERPO

- **PR** \( \min_{\theta, q} \mathcal{L}(\theta, q) = KL(q(x) || p_\theta(x)) - \alpha \mathbb{E}_q [f_\phi(x)] \)

- **ERPO** \( \min_{q_\pi} \mathcal{L}(q_\pi) = KL(q_\pi(x) || p_\pi(x)) - \alpha \mathbb{E}_q [R(x)] \)

**PR**
- Generative model \( p_\theta(x) \)
- Constraint \( f_\phi(x) \)

**ERPO**
- Reference (old) policy \( p_\pi(x) \)
- Reward \( R(x) \)

\[ q^*_\phi(y|x) \propto p_\theta(x) \exp\{ \alpha f_\phi(x) \} \]

\[ q^*_\pi(y|x) \propto p_\pi(x) \exp\{ \alpha R(x) \} \]
Close resemblance b/w PR and ERPO

- **PR** \( \min_{\theta, q} \mathcal{L}(\theta, q) = \text{KL}(q(x) || p_\theta(x)) - \alpha \mathbb{E}_q[f_\phi(x)] \)

- **ERPO** \( \min_{q_\pi} \mathcal{L}(q_\pi) = \text{KL}(q_\pi(x) || p_\pi(x)) - \alpha \mathbb{E}_q[R(x)] \)

<table>
<thead>
<tr>
<th>PR</th>
<th>ERPO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generative model ( p_\theta(x) )</td>
<td>Reference (old) policy ( p_\pi(x) )</td>
</tr>
<tr>
<td>Constraint ( f_\phi(x) )</td>
<td>Reward ( R(x) )</td>
</tr>
</tbody>
</table>

\[ q_\phi^*(y|x) \propto p_\theta(x) \exp\{ \alpha f_\phi(x) \} \leftrightarrow q_\pi^*(y|x) \propto p_\pi(x) \exp\{ \alpha R(x) \} \]

Inverse reinforcement learning (IRL) can learn a reward function from data.
Close resemblance b/w PR and ERPO

- PR \[ \min_{\theta, q} \mathcal{L}(\theta, q) = KL(q(\mathbf{x}) \| p_\theta(\mathbf{x})) - \alpha \mathbb{E}_q[f_\phi(\mathbf{x})] \]

- ERPO \[ \min_{q_\pi} \mathcal{L}(q_\pi) = KL(q_\pi(\mathbf{x}) \| p_\pi(\mathbf{x})) - \alpha \mathbb{E}_q[R(\mathbf{x})] \]

Inverse reinforcement learning (IRL) can learn a reward function from data

We can transfer IRL technique to learn \( f_\phi(\mathbf{x}) \)!
Maximum-Entropy Inverse Reinforcement Learning (MaxEnt IRL)

• ERPO  \( \min_{q_\pi} L(q_\pi) = \text{KL}(q_\pi(x)||p_\pi(x)) - \alpha \mathbb{E}_q[R(x)] \)

\[ q_{\pi}^*(y|x) \propto p_\pi(x) \exp\{ \alpha R(x) \} \]

• MaxEnt IRL learns reward function \( R_\phi(x) \) with unknown parameters \( \phi \)

• Assumes reference policy \( p_\pi(x) \) as a uniform, so the above KL regularization becomes an entropy regularization (i.e., MaxEnt)

• The above \( q_{\pi}^*(y|x) \) now additionally depends on \( \phi \)

\[ q_{\pi,\phi}(y|x) \propto \exp\{ \alpha R_\phi(x) \} \]

• Learns \( \phi \) by maximizing data log-likelihood

\[ \phi^* = \text{argmax}_\phi \mathbb{E}_{x \sim p_{\text{data}}} [\log q_{\pi,\phi}(x)] \]
Maximum-Entropy Inverse Reinforcement Learning (MaxEnt IRL)

• ERPO \( \min_{q_\pi} \mathcal{L}(q_\pi) = \text{KL}(q_\pi(x)||p_\pi(x)) - \alpha \mathbb{E}_q[R(x)] \)

  \[ q_\pi^*(y|x) \propto p_\pi(x) \exp\{\alpha R(x)\} \]

• MaxEnt IRL learns reward function \( R_\phi(x) \) with unknown parameters \( \phi \)

• Assumes reference policy \( p_\pi(x) \) as a uniform, so the above KL regularization becomes an entropy regularization (i.e., MaxEnt)

• The above \( q_\pi^*(y|x) \) now additionally depends on \( \phi \)

  \[ q_{\pi,\phi}(y|x) \propto \exp\{\alpha R_\phi(x)\} \]

• Learns \( \phi \) by maximizing data log-likelihood

  \[ \phi^* = \arg\max_{\phi} \mathbb{E}_{x\sim p_{\text{data}}} [\log q_{\pi,\phi}(x)] \]

Apply this objective to learn \( f_\phi \) in PR
Algorithm: PR with learnable constraint

- **PR:** \( \min_{\theta,q} \mathcal{L}(\theta, q) = \text{KL}(q(x)\|p_{\theta}(x)) - \alpha \mathbb{E}_q[f_{\phi}(x)] \)

- **(1) Learning the constraint** \( f_{\phi} \)
  
  **E-step:** \( q_{\phi}(y|x) \propto p_{\theta}(x)\exp\{\alpha f_{\phi}(x)\} \)
  
  - Use the same objective as in MaxEnt IRL
  
  \( \phi^* = \arg\max_{\phi} \mathbb{E}_{x \sim p_{\text{data}}} [\log q_{\phi}(x)] \)

- **(2) Learning the generative model** \( p_{\theta} \)
  
  **M-step:** \( \min_{\theta} \text{KL}(q_{\phi}(x)\|p_{\theta}(x)) = \min_{\theta} -\mathbb{E}_{q_{\phi}}[\log p_{\theta}(x)] + \text{const.} \)
  
  - If \( p_{\theta} \) is an implicit model (e.g., GANs), we can approximate by minimizing reverse KL \( \min_{\theta} \text{KL}(p_{\theta}(x)\|q_{\phi}(x)) \)
Experiments – Pose-conditional Human Image Generation

- Source image
- Generative model $p_\theta$
- Human part parser
- Learnable module $\phi$
- Constraint
- Structured consistency
Experiments – Pose-conditional Human Image Generation

<table>
<thead>
<tr>
<th>Method</th>
<th>SSIM</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Ma et al. [38]</td>
<td>0.614</td>
<td>—</td>
</tr>
<tr>
<td>2 Pumarola et al. [44]</td>
<td>0.747</td>
<td>—</td>
</tr>
<tr>
<td>3 Ma et al. [37]</td>
<td>0.762</td>
<td>—</td>
</tr>
<tr>
<td>4 Base model</td>
<td>0.676</td>
<td>0.03</td>
</tr>
<tr>
<td>5 With fixed constraint</td>
<td>0.679</td>
<td>0.12</td>
</tr>
<tr>
<td>6 With learned constraint</td>
<td><strong>0.727</strong></td>
<td><strong>0.77</strong></td>
</tr>
</tbody>
</table>

Results of image generation on Structural Similarity (SSIM) between generated and true images, and human survey where the full model yields better generations than the base models (Rows 5-6) on 77% test cases.
Experiments – Pose-conditional Human Image Generation

<table>
<thead>
<tr>
<th>source image</th>
<th>target pose</th>
<th>target image</th>
<th>Learned constraint</th>
<th>Fixed constraint</th>
<th>Base model</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Image]</td>
<td>[Image]</td>
<td>[Image]</td>
<td>[Image]</td>
<td>[Image]</td>
<td>[Image]</td>
</tr>
<tr>
<td>[Image]</td>
<td>[Image]</td>
<td>[Image]</td>
<td>[Image]</td>
<td>[Image]</td>
<td>[Image]</td>
</tr>
</tbody>
</table>

Samples generated by the models. The model with learned human part constraint generates correct poses and preserves human body structure much better.
Experiments – Template-guided Sentence Generation

- **Task:** Given a template, generate a complete sentence following the template
- **Constraint:** force the match between the *infilling content* of the generated sentence with the true content

**template:**

"____ meant to ____ not to ______"

**true target:**

"It was meant to dazzle not to make sense."

**generated:**

"It was meant to dazzle not to make it."

---

**Constraint $f_\phi$:**

Learnable module $\phi$

Infilling content matching
Experiments – Template-guided Sentence Generation

<table>
<thead>
<tr>
<th>Model</th>
<th>Perplexity</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Base model</td>
<td>30.30</td>
<td>0.19</td>
</tr>
<tr>
<td>2 With binary D</td>
<td>30.01</td>
<td>0.20</td>
</tr>
<tr>
<td>3 With constraint updated in M-step (Eq.5)</td>
<td>31.27</td>
<td>0.15</td>
</tr>
<tr>
<td>4 With learned constraint</td>
<td>28.69</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Samples by the full model are considered as of higher quality in 24% cases.

Two test examples, including the template, the sample by the base model, and the sample by the constrained model.

```
_ the_ acting
is the acting .

_ the_ acting
is also very good .

I will give the movie 7

out of 10 .

10

out of 10 .
```
Conclusions

• Formal connections between posterior regularization (PR) and reinforcement learning (RL)

• Learn the knowledge constraints in PR as reward learning in (inverse) RL

• The resulting algorithm is:
  • generally applicable to any deep generative models
  • flexible to learn the constraints and model jointly

• Experiments on image and text generation showed the effectiveness of the algorithm