JointGAN: Multi-Domain Joint Distribution Learning with Generative Adversarial Nets

Presenter: Shuyang Dai

Yunchen Pu, Shuyang Dai, Zhe Gan, Weiyao Wang, Guoyin Wang, Yizhe Zhang, Ricardo Henao, Lawrence Carin

Facebook & Duke University & Microsoft Research
Joint Distribution Learning

• Learning *marginals* with GAN:

$$\tilde{x} = f_\alpha(\epsilon_1), \quad \epsilon_1 \sim p(\epsilon_1)$$
$$\tilde{y} = f_\beta(\epsilon'_1), \quad \epsilon'_1 \sim p(\epsilon'_1)$$

• Learning *conditionals* with ALI, Cycle GAN, and Triangle GAN:

$$\tilde{y} = f_\theta(x, \epsilon_2), \quad x \sim q(x), \quad \epsilon_2 \sim p(\epsilon_2)$$
$$\tilde{x} = f_\phi(y, \epsilon'_2), \quad y \sim q(y), \quad \epsilon'_2 \sim p(\epsilon'_2)$$
Though joint distributions are matched, only conditional distributions are learned.

Is there a way to fully learn the joint distribution?
Joint Distribution Learning

• Potential advantages:
  • Synthesis of draws from any of the marginals
  • Synthesis of draws from the conditionals when other random variables are observed, \textit{i.e.}, imputation
  • Also allows complete draws from the full joint distribution
Joint Distribution Learning

• Learning joint distribution with the proposed JointGAN:

\[
\begin{align*}
    x &= f_\alpha(\epsilon_1) \\
    y &= f_\theta(x, \epsilon_2)
\end{align*}
\]

\[
\begin{align*}
    \epsilon_1 &\sim p(\epsilon_1) \\
    \epsilon_2 &\sim p(\epsilon_2)
\end{align*}
\]

\[
\begin{align*}
    x &= f_\phi(y, \epsilon'_2) \\
    y &= f_\beta(\epsilon'_1)
\end{align*}
\]

\[
\begin{align*}
    \epsilon'_1 &\sim p(\epsilon'_1) \\
    \epsilon'_2 &\sim p(\epsilon'_2)
\end{align*}
\]

\[
(x, y) \sim p_\alpha(x)p_\theta(y|x) \quad (x, y) \sim p_\beta(y)p_\phi(x|y)
\]

• To reduce the number of parameters, let \( f_\alpha(\cdot) = f_\phi(0, \cdot) \) and \( f_\beta(\cdot) = f_\theta(0, \cdot) \), where 0 is an all-zero tensor.
Training JointGAN

\[ \begin{align*}
    \epsilon_1, \epsilon'_1 \quad &\xrightarrow{x \sim q(x)} \quad y = f_\theta(x, \epsilon_3) \\
    x = f_\phi(0, \epsilon_1) \quad &\xrightarrow{y = f_\theta(x, \epsilon_2)} \\
    y = f_\theta(0, \epsilon'_1) \quad &\xrightarrow{x = f_\phi(y, \epsilon'_2)} \\
    y \sim q(y) \quad &\xrightarrow{x = f_\phi(y, \epsilon'_3)} \\
    x, y \sim q(x, y) \quad &\xrightarrow{\text{Fake Pairs}} \\
    x, y \sim q(x, y) \quad &\xrightarrow{\text{Real Pairs}} 
\end{align*} \]
Training JointGAN

• A single 5-way critic (discriminator):

\[
p_1(x, y) = q(x)p_\theta(y|x), \quad p_2(x, y) = q(y)p_\phi(x|y) \\
\]

\[
p_3(x, y) = p_\alpha(x)p_\theta(y|x), \quad p_4(x, y) = p_\beta(y)p_\phi(x|y) \\
\]

\[
p_5(x, y) = q(x, y) 
\]

(3)

• The minimax objective for JointGAN:

\[
\min_{\theta, \phi} \max_{\omega} \mathcal{L}_{\text{JointGAN}}(\theta, \phi, \omega) \\
= \sum_{k=1}^{5} \mathbb{E}_{p_k(x, y)}[\log g_\omega(x, y)[k]] 
\]

(4)

where the critic \( g_\omega(x, y) \in \Delta^4 \) has softmax on the top layer:

\[
\sum_{k=1}^{5} g_\omega(x, y)[k] = 1 \text{ and } g_\omega(x, y)[k] \in (0, 1) 
\]

(5)
JointGAN: Computational Cost

• The capacity of our model is increased without introducing additional computational cost

• The parameters of the generators are tied together:
  
  - $f_\alpha(\cdot) = f_\phi(0, \cdot)$ and $f_\beta(\cdot) = f_\theta(0, \cdot)$
  - Almost no additional parameters when compared with ALI

• A single 5-way critic (discriminator):
  
  - Also, almost no additional parameters when compared with ALI
Experiment: generated image pairs

Figure 1. Generated paired samples from models trained on paired data.
**Figure 2.** Generated paired samples of caption features and images. Left block: from generated images to caption features. Right block: from generated caption features to images.
Handle Discrete Text Data

- sentence → encoder → latent feature → decoder → reconstruction
- $\epsilon$ → generator → latent feature
- discriminantor
- real / fake → generation
Extension I: No Paired Data

• When paired draws from $p_5(x, y) = q(x, y)$ are not available, cycle consistency is imposed:

$$\min_{\theta, \phi} \max_\omega \mathcal{L}_{\text{JointGAN}}(\theta, \phi, \omega) = \sum_{k=1}^4 \mathbb{E}_{p_k(x, y)}[\log g'_\omega(x, y)[k]] + R_{\theta, \phi}(x, y)$$

(6)

where $R_{\theta, \phi}(x, y)$ is a regularization term:

$$R_{\theta, \phi}(x, y) = \mathbb{E}_{x \sim q(x), y \sim p_\theta(y|x), \hat{x} \sim p_\phi(x|y)} ||x - \hat{x}|| + \mathbb{E}_{y \sim q(y), x \sim p_\theta(x|y), \hat{y} \sim p_\phi(y|x)} ||y - \hat{y}||$$

(7)
Extension I: No Paired Data

Figure 3. Generated paired samples from models trained on unpaired data.
Extension II: Multiple Domains

• Consider two specific forms for joint random variables \((x, y, z)\):

\[
p(x, y, z) = p_\alpha(x)p_\nu(y|x)p_\gamma(z|x, y) = p_\beta(z)p_\psi(y|z)p_\eta(x|y, z)
\]  

(8)

• Assume only empirical draws from \(q(x, y)\) and \(q(y, z)\) are available. Let the critic be a 6-way softmax classifier:

\[
\begin{align*}
p_\alpha(x)p_\nu(y|x)p_\gamma(z|x, y), & \quad p_\beta(z)p_\psi(y|z)p_\eta(x|y, z) \\
q(x)p_\nu(y|x)p_\gamma(z|x, y), & \quad q(x)p_\psi(y|z)p_\eta(x|y, z) \\
q(x, y)p_\gamma(z|x, y), & \quad q(y, z)p_\eta(x|y, z).
\end{align*}
\]

(9)
Figure 4. Generated paired samples from models trained on facades↔labels↔cityscapes. Left: sequentially generated from left to right. Right: generated from right to left.
Quantitative results

• Compare with a two-step baseline:
  • WGAN-GP is employed to model the two marginals
  • Pix2pix and CycleGAN are utilized to model the conditionals for the case w/wo paired empirical draws

<table>
<thead>
<tr>
<th>Table 1. Human evaluation results on the quality of generated pairs.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Method</strong></td>
</tr>
<tr>
<td><strong>Trained with paired data</strong></td>
</tr>
<tr>
<td>WGAN-GP + Pix2pix wins</td>
</tr>
<tr>
<td>JointGAN wins</td>
</tr>
<tr>
<td>Not distinguishable</td>
</tr>
<tr>
<td><strong>Trained with unpaired data</strong></td>
</tr>
<tr>
<td>WGAN-GP + CycleGAN wins</td>
</tr>
<tr>
<td>JointGAN wins</td>
</tr>
<tr>
<td>Not distinguishable</td>
</tr>
</tbody>
</table>
Quantitative results

• We further define **Relevance Score**
  • Used to evaluate the quality and relevance of two generated images
  • Calculated as the cosine similarity between two images that are embedded into a shared latent space, which are learned via training a ranking model

<table>
<thead>
<tr>
<th></th>
<th>edges↔shoes</th>
<th>edges↔handbags</th>
<th>labels↔cityscapes</th>
<th>labels↔facades</th>
<th>maps↔satellites</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>True pairs</strong></td>
<td>0.684</td>
<td>0.672</td>
<td>0.591</td>
<td>0.529</td>
<td>0.514</td>
</tr>
<tr>
<td><strong>Random pairs</strong></td>
<td>0.008</td>
<td>0.005</td>
<td>0.012</td>
<td>0.011</td>
<td>0.054</td>
</tr>
<tr>
<td><strong>Other pairs</strong></td>
<td>0.113</td>
<td>0.139</td>
<td>0.092</td>
<td>0.076</td>
<td>0.081</td>
</tr>
<tr>
<td><strong>WGAN-GP + Pix2pix</strong></td>
<td>0.352</td>
<td>0.343</td>
<td>0.301</td>
<td>0.288</td>
<td>0.125</td>
</tr>
<tr>
<td><strong>JointGAN (paired)</strong></td>
<td>0.488</td>
<td>0.489</td>
<td>0.377</td>
<td>0.364</td>
<td>0.328</td>
</tr>
<tr>
<td><strong>WGAN-GP + CycleGAN</strong></td>
<td>0.203</td>
<td>0.195</td>
<td>0.201</td>
<td>0.139</td>
<td>0.091</td>
</tr>
<tr>
<td><strong>JointGAN (unpaired)</strong></td>
<td>0.452</td>
<td>0.461</td>
<td>0.339</td>
<td>0.341</td>
<td>0.299</td>
</tr>
</tbody>
</table>
Summary

• We propose JointGAN for multi-domain joint distribution learning

• JointGAN provides freedom to draw samples from various marginalized or conditional distributions

• Besides image pairs, we also provided generated image-caption pairs
Come to our poster for details
@ Hall B #109

Yunchen Pu, Shuyang Dai, Zhe Gan, Weiyao Wang, Guoyin Wang, Yizhe Zhang, Ricardo Henao, Lawrence Carin

Facebook & Duke University & Microsoft Research