Graph Convolutional Policy Network for Goal-Directed Molecular Graph Generation

Jiaxuan You      Bowen Liu
Rex Ying      Vijay Pande      Jure Leskovec

Presenter: Yevgeny Tkach

2019 Spring @
https://qdata.github.io/deep2Read/
Executive Summary

- Molecule generation using GAN approach
- Edge-by-edge generation based on reinforcement learning
- Loss function gives is a weighted sum of adversarial loss, domain specific target metrics, validity constraints
- Target metrics are optimized during the learning and not as a post learning optimization process which offers state of the art results in the domain
- Graph representation is performed with GCN
- Training of the generator performed with policy gradient
Graph $G = (A, E, F)$, $A \in \{0, 1\}^{n \times n}$, $F \in \mathbb{R}^{n \times d}$, $E \in \{0, 1\}^{b \times n \times n}$, $A = \sum_{i=1}^{b} E_i$

- Run GCN
  $H^{(l+1)} = \text{AGG} (\text{ReLU} (\{ \tilde{D}^{-\frac{1}{2}} \tilde{E} \tilde{D}^{-\frac{1}{2}} H^{(l)} W_i^{(l)} \}, \forall i \in (1, \ldots, b)))$
GCPN

- Run 4 MLP networks to build a 4 part action:

\[ a_t = \text{CONCAT}(a_{\text{first}}, a_{\text{second}}, a_{\text{edge}}, a_{\text{stop}}) \]

\[
\begin{align*}
    f_{\text{first}}(s_t) &= \text{SOFTMAX}(m_f(X)), \\
    f_{\text{second}}(s_t) &= \text{SOFTMAX}(m_s(X_{a_{\text{first}}}, X)), \\
    f_{\text{edge}}(s_t) &= \text{SOFTMAX}(m_e(X_{a_{\text{first}}}, X_{a_{\text{second}}})), \\
    f_{\text{stop}}(s_t) &= \text{SOFTMAX}(m_t(\text{AGG}(X))), \\
    a_{\text{first}} &\sim f_{\text{first}}(s_t) \in \{0, 1\}^n, \\
    a_{\text{second}} &\sim f_{\text{second}}(s_t) \in \{0, 1\}^{n+c}, \\
    a_{\text{edge}} &\sim f_{\text{edge}}(s_t) \in \{0, 1\}^b, \\
    a_{\text{stop}} &\sim f_{\text{stop}}(s_t) \in \{0, 1\}
\end{align*}
\]
Check whether the suggested action is legal, based on a 3rd party molecule checker, specifically RDKit.

If it is legal update the state if it is not resemble from the policy and repeat until a legal action is proposed.

Keep track of the “step rewards” which adds and removes punishments for suggesting legal or illegal actions.
**Reward function**

- The reward function is built from intermediate/step reward and final reward.

- Final reward is a weighted sum over domain specific rewards and an adversarial reward.

- Domain specific rewards include scores like octanol-water partition coefficient (logP), druglikeness (QED) and molecular weight (MW).

- Adversarial reward is computed by training a discriminator using a GAN loss to distinguish between generated and true molecules. The discriminator is also a GCN.

\[
\min_{\theta} \max_{\phi} V(\pi_\theta, D_\phi) = \mathbb{E}_{x \sim p_{data}} [\log D_\phi(x)] + \mathbb{E}_{x \sim \pi_\theta} [\log D_\phi(1 - x)]
\]
Experiments

- **Property Optimization** - The task is to generate novel molecules whose specified molecular properties are optimized.

- **Property Targeting** - The task is to generate novel molecules whose specified molecular properties are as close to the target scores as possible.

- **Constrained Property Optimization** - The task is to generate novel molecules whose specified molecular properties are optimized, while also containing a specified molecular substructure.
## Experiment results

<table>
<thead>
<tr>
<th>Method</th>
<th>Penalized logP</th>
<th>QED</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st</td>
<td>2nd</td>
<td>3rd</td>
<td>Validity</td>
<td>1st</td>
<td>2nd</td>
</tr>
<tr>
<td>ZINC</td>
<td>4.52</td>
<td>4.30</td>
<td>4.23</td>
<td>100.0%</td>
<td>0.948</td>
<td>0.948</td>
</tr>
<tr>
<td>Hill Climbing</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td></td>
<td>0.838</td>
<td>0.814</td>
</tr>
<tr>
<td>ORGAN</td>
<td>3.63</td>
<td>3.49</td>
<td>3.44</td>
<td>0.4%</td>
<td>0.896</td>
<td>0.824</td>
</tr>
<tr>
<td>JT-VAE</td>
<td>5.30</td>
<td>4.93</td>
<td>4.49</td>
<td>100.0%</td>
<td>0.925</td>
<td>0.911</td>
</tr>
<tr>
<td>GCPN</td>
<td>7.98</td>
<td>7.85</td>
<td>7.80</td>
<td>100.0%</td>
<td>0.948</td>
<td>0.947</td>
</tr>
</tbody>
</table>
### Table 2: Comparison of the effectiveness of property targeting task.

<table>
<thead>
<tr>
<th>Method</th>
<th>$-2.5 \leq \log P \leq -2$</th>
<th>Success</th>
<th>Diversity</th>
<th>$5 \leq \log P \leq 5.5$</th>
<th>Success</th>
<th>Diversity</th>
<th>$150 \leq MW \leq 200$</th>
<th>Success</th>
<th>Diversity</th>
<th>$500 \leq MW \leq 550$</th>
<th>Success</th>
<th>Diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZINC</td>
<td>0.3%</td>
<td>0.919</td>
<td>1.3%</td>
<td>0.909</td>
<td>1.7%</td>
<td>0.938</td>
<td>0</td>
<td>–</td>
<td>–</td>
<td>16.0%</td>
<td>0.898</td>
<td>–</td>
</tr>
<tr>
<td>JT-VAE</td>
<td>11.3%</td>
<td>0.846</td>
<td>7.6%</td>
<td>0.907</td>
<td>0.7%</td>
<td>0.824</td>
<td>16.0%</td>
<td>0.898</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORGAN</td>
<td>0</td>
<td>–</td>
<td>0.2%</td>
<td>0.909</td>
<td>15.1%</td>
<td>0.759</td>
<td>0.1%</td>
<td>0.907</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCPN</td>
<td>85.5%</td>
<td>0.392</td>
<td>54.7%</td>
<td>0.855</td>
<td>76.1%</td>
<td>0.921</td>
<td>74.1%</td>
<td>0.920</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 3: Comparison of the performance in the constrained optimization task.

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>JT-VAE</th>
<th>GCPN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Improvement</td>
<td>Similarity</td>
</tr>
<tr>
<td>0.0</td>
<td>$1.91 \pm 2.04$</td>
<td>$0.28 \pm 0.15$</td>
</tr>
<tr>
<td>0.2</td>
<td>$1.68 \pm 1.85$</td>
<td>$0.33 \pm 0.13$</td>
</tr>
<tr>
<td>0.4</td>
<td>$0.84 \pm 1.45$</td>
<td>$0.51 \pm 0.10$</td>
</tr>
<tr>
<td>0.6</td>
<td>$0.21 \pm 0.71$</td>
<td>$0.69 \pm 0.06$</td>
</tr>
</tbody>
</table>
Discussion

• The authors present a very clean policy training approach without sophisticated choice of vocabulary.
• Junctional tree approach was not necessarily useful in molecule generation - It would be interesting to check the scaffold approach with RL training framework.
• Legality check of actions during training and testing is a powerful technic.
• Where applicable RL approach to optimize a target metric seems to be more powerful then a gradient based optimization of an approximate function.