Improving Small Molecule Generation using Mutual Information Machine

Danny Reidenbach, Micha Livne, Rajesh K. Ilango, Michelle Gill, Johnny Israeli (contact: Micha Livne mlivne@nvidia.com)

Background

MolMIM is a probabilistic auto-encoder for small molecules, pretrained on 730M SMILES molecules randomly selected from drug-like tranches of ZINC15. Mutual Information Machine (MIM) learning maximizes the mutual information between the observations and latent codes, while clustering similar observations together in the latent space.

Algorithm 1: Learning parameters $\boldsymbol{\theta}$ of MolMIM

Require: Samples from dataset $\mathcal{P}(\boldsymbol{x})$

while not converged **do**

$\sigma \sim \mathcal{U}(0, 1]$

$D \leftarrow \{\boldsymbol{x}_j, \boldsymbol{z}_j \sim q_{\boldsymbol{\theta}}(\boldsymbol{z}|\boldsymbol{x}, \sigma)\mathcal{P}(\boldsymbol{x})\}_{j=1}^{N}$

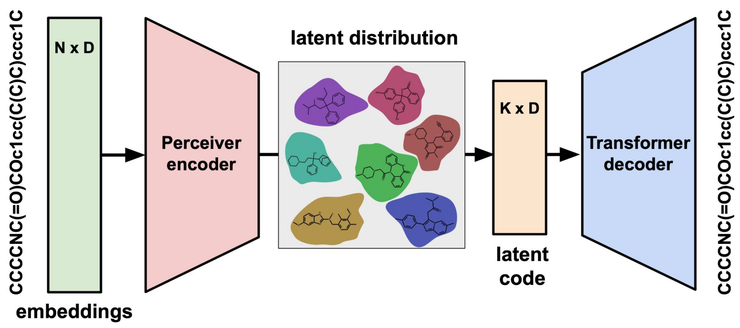
$\hat{\mathcal{L}}_{\text{A-MIM}}(\boldsymbol{\theta}; D) = -\frac{1}{N}\sum_{i=1}^{N}\left(\log p_{\boldsymbol{\theta}}(\boldsymbol{x}_i|\boldsymbol{z}_i) + \frac{1}{2}\left(\log q_{\boldsymbol{\theta}}(\boldsymbol{z}_i|\boldsymbol{x}_i, \sigma) + \log \mathcal{P}(\boldsymbol{z}_i)\right) \right)$

$\Delta\boldsymbol{\theta} \propto -\nabla_{\boldsymbol{\theta}}\hat{\mathcal{L}}_{\text{A-MIM}}(\boldsymbol{\theta}; D)$ {*Gradient computed through sampling using reparameterization*}

end while

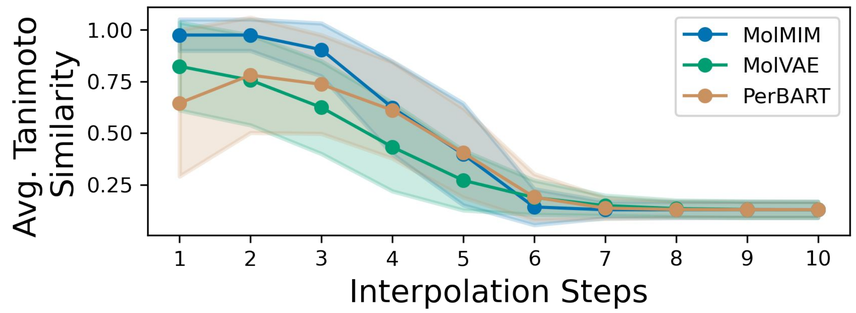
We show how MolMIM's learned latent space can be used with a naïve CMA-ES optimization algorithm to surpass a wide array of complex generative methods in drug discovery tasks.

Architecture



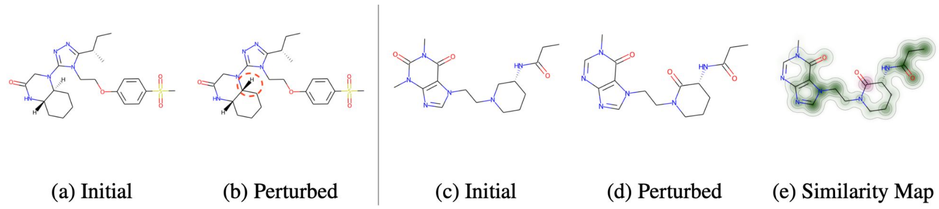
MolMIM is an encoder-decoder with a fixed-size bottleneck, built using a Perceiver encoder and Transformer decoder. While similar in architecture to β-VAE, MolMIM is a Mutual Information Machine (MIM), and does not suffer from posterior collapse.

Experiment: Clustering in the Latent Space



Average Tanimoto similarity (y-axis) between input and non-identical interpolated molecules (x-axis is interpolation step in the latent code). Shaded regions indicate +/- one standard deviation. MolMIM clusters similar molecules and has the smallest variance in early steps. The clustering lends itself to various tasks of controlled generation.

Experiment: Fine-grained Control of Generation



(a) Initial (b) Perturbed (c) Initial (d) Perturbed (e) Similarity Map

MolMIM provides fine-grained control over molecule generation: **(a-b)** Small perturbations lead to changes in chirality only (dashed red line); **(c-d)** Larger changes allow the substitution of a single atom, see similarity map (e); **(e)** The similarity map with chemically similar (green) and distinct (red) pharmacophores, relative to the initial structure (c).

Experiment: Sampling

Model	K	Latent Dim.	Eff. Nov.(%)	Validity(%)	Unique(%)	Non Id.(%)	Novelty(%)	σ	Test Time	Batch
MMB	-	variable	51.1	75	84.8	74.4	93.1	1.2	8.7 hours	100 †
PerBART	4	2048	59.1	71.8	94.9	88.4	94.3	0.7	38 min	500
MolVAE	4	2048	93.9	95.7	**100**	**100**	98.1	1.2	63 min	500
MolMIM	1	**512**	**94.2**	**98.7**	**100**	99.9	95.5	1.42	**30 min**	500
CDDD	1	**512**	82.2	84.5	98.9	98	**99.4**	1.2	12 hours	1

MolMIM archives the highest effective novelty (i.e., valid, novel, and unique molecules) in the shortest amount of inference time. Molecule sampling quality was evaluated with 20,000 molecules randomly selected from the test set, where 10 samples were acquired per molecule. MMB stands for MegaMolBART. K is the hidden length. σ is the optimal scale of Gaussian random noise used in sampling. † batch size constrained by memory.

Experiment: Controlled Generation

Task	QED (%) $\delta = 0.4$	Penalized logP $\delta = 0.6$	Penalized logP $\delta = 0.4$
JT-VAE	8.8	0.28 ± 0.79	1.03 ± 1.39
GCPN	9.4	0.79 ± 0.63	2.49 ± 1.30
MolDQN	-	1.86 ± 1.21	3.37 ± 1.62
MMPA	32.9	-	-
VSeq2Seq	58.5	2.33 ± 1.17	3.37 ± 1.75
VJTNN+GAN	60.6	-	-
VJTNN	-	2.33 ± 1.24	3.55 ± 1.67
MoFlow	-	2.10 ± 2.86	4.71 ± 4.55
GA	-	3.44 ± 1.09	5.93 ± 1.41
AtomG2G	73.6	-	-
HierG2G	76.9	-	-
DESMILES	77.8	-	-
QMO	92.8	3.73 ± 2.85	7.71 ± 5.65
MolGrow	-	4.06 ± 5.61	8.34 ± 6.85
GraphAF	-	4.98 ± 6.49	8.21 ± 6.51
GraphDF	-	4.51 ± 5.80	9.19 ± 6.43
CDGS	-	5.10 ± 5.80	9.56 ± 6.33
FaST	-	8.98 ± 6.31	18.09 ± 8.72
MolMIM	**94.6**	$\mathbf{7.60 \pm 23.62}$	$\mathbf{28.45 \pm 54.67}$
MolMIM†		4.57 ± 3.87	9.44 ± 4.12

QED and penalized logP optimization under minimal Tanimoto similarity constraint δ. MolMIM performs better in constrained optimization than models trained with supervision. QED is quantitated by success percentage and penalized logP by mean and standard deviation of the improvement in value. † limits logP solutions to improvement ≤ 20 (removing long carbon chains), Results above the solid bar are from [Fragment-Based Sequential Translation for Molecular Optimization, Chen, 2021].

Model	GSK3β + JNK3 + QED + SA Success (%)	Novelty (%)	Diversity
JT-VAE	1.3	-	-
GVAE-RL	2.1	-	-
GCPN	4.0	-	-
REINVENT	47.9	-	-
RationaleRL	74.8	56.1	0.621
MARS	92.3	82.4	0.719
JANUS	**100**	32.6	**.821**
FaST	**100**	**100**	0.716
MolMIM (R)	97.5	71.1	0.791
MolMIM (A)	96.6	63.3	0.807
MolMIM (E)	98.3	55.1	0.767
MolMIM (E)†	99.2	54.8	0.772

Multi-objective molecule optimization. MolMIM is comparable to SOTA models which were specifically designed for the multi-objective task.The MolMIM experiments are denoted as: **(R)** random initialization; **(A)** promising precursor initialization; **(E)** initialization with exemplars. **†** results are based on additional restarts