16

Review of Quantum Generative Models for Small Molecule Drug Discovery

Wali Mirza

Indiana University Bloomington

wmirza@iu.edu

Introduction

Background:

- Drug discovery is a lengthy and expensive process, often taking over a decade and costing billions of dollars. This involves searching a vast molecular "chemical space," estimated to have up to 10^{60} possibilities.
- Machine learning models, like GANs, have been applied to this problem, successfully generating molecules with desired chemical and physical properties.

Challenges:

- Classical GANs struggle with the vast chemical space due to limitations in computational capacity and the curse of dimensionality.
- A fully quantum GAN would require over 90 qubits for small molecule datasets like QM9, exceeding current quantum computer capabilities.

Proposed Solution:

- The QGAN-HG (Hybrid Quantum GAN) combines the strengths of quantum computing and classical neural networks:
  1. Quantum Generator: Uses qubits to explore the chemical space efficiently, leveraging quantum parallelism.
  2. Classical Discriminator: Evaluates and refines generated molecules to align with real data distributions.
- This approach reduces the parameters needed while maintaining or improving performance compared to classical models.

Objective

Develop a qubit-efficient hybrid quantum GAN (QGAN-HG) that combines quantum and classical methods to efficiently generate valid drug-like molecules, overcoming the limitations of classical GANs in exploring vast chemical spaces.

Methods

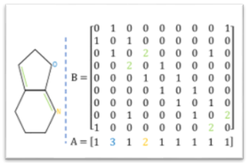
Hybrid Quantum-Classical Architecture:

- The **generator** consists of:
  1. A **quantum circuit** that creates a feature vector by leveraging qubit superposition and entanglement to explore the chemical space efficiently.
  2. A **classical neural network** that transforms the quantum-generated features into atom vectors and bond matrices, representing drug molecules as graphs.
- The **discriminator** is a classical neural network that distinguishes between real molecules from the dataset and synthetic molecules generated by the model.

Molecular Representation:

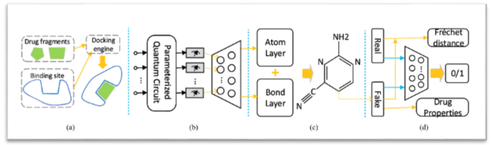
Molecules are encoded as:

- **Atom vectors**: Each entry describes atom types.
- **Bond matrices**: Square matrices with entries representing bond types and properties.



Workflow:

1. **Input Noise**: Quantum circuit takes uniformly sampled noise (e.g., from [−π,π]x[-π, π]) to generate a random quantum state.
2. **Feature Generation**: Quantum circuit processes noise to create a feature superposition using parameterized gates, then the expected value of each feature is approximated and used in the classical neural network.
3. **Molecule Generation**: Classical neural network outputs atom and bond representations.
4. **Validation**: Discriminator evaluates the generated molecules, feeding back error signals to improve the generator.



Training:

- The model optimizes Fréchet Distance (FD) to align generated molecule distributions with real data. Training 85% reduced (Medium Reduced/MR) MolGAN models to measure the effect with and without a hybrid quantum generator.
- Modifications to reduce model instability. Changed parameter initialization, learning rate strategy and quantum circuit design flaws. Experiment reran 8 times for more certainty on results.

Results

Performance Metrics:

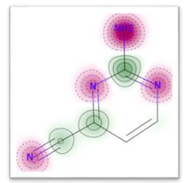
- **Fréchet Distance (FD):** Measures how closely the generated molecule distribution matches the real molecule distribution.
  - QGAN-HG achieved a **lower FD** compared to reduced classical GANs, indicating superior alignment with the real data distribution.
- **Molecule Validity:** Generated molecules were consistently valid, with QGAN-HG outperforming MolGAN in valid score by **~30(/100) points on average**.

Models\Metrics	FD Score	Valid Score	Druglikeliness	Solubility	Synthesizability
MolGAN (MR)	11.42 (± 1.08)	46.71 (± 13.18)	0.47 (± 0.00)	0.66 (± 0.03)	0.12(± 0.05)
QGAN-HG (MR) 10Q	10.71 (± 1.62)	**76.64 (± 9.74)**	0.49 (± 0.01)	0.67 (± 0.05)	0.10 (± 0.06)

Drug Property Evaluation:

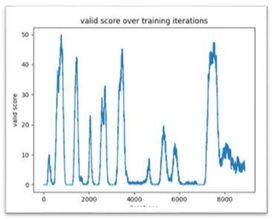
- Metrics like drug-likeness (QED), solubility (logP), and synthesizability (SA) were very similar with and without quantum circuit.
- QGAN-HG with 85% less parameters had a synthesizability score of **~70(/100) points lower on average** than the original MolGAN.

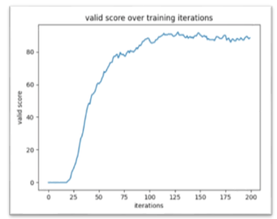
Generated Molecule Example:



Stability and Training:

- Code modifications increased model stability and performance. Converges in substantially less iterations.





Discussion and Future Work

Model Advantages:

- The hybrid quantum-classical architecture demonstrated strong potential for efficiently exploring the vast chemical space. The lower average Fréchet Distance and significantly higher molecule validity scores validate the effectiveness of the QGAN-HG design for molecular graph generation.

Limitations:

- **Instability in Training**: model collapse occurred more often with the quantum circuit included, however this should be less of problem using different GAN variations that address the inherit instability in its design.
- **Scalability Concerns**: Current implementations are limited to small molecules due to qubit constraints and noise in near-term quantum devices. There isn't proof that scaling qubits will retain its significance to bigger datasets.

Comparison with Classical Models:

- While QGAN-HG outperformed reduced classical GANs, full MolGAN models still exceeded QGAN-HG on some metrics like synthesizability.

Future Work:

1. **Training Instability:** Explore variations of the GAN, like the variational autoencoder that are more stable. Test results on real quantum computers
2. **Qubit Scalability:** Incorporate more qubits and leverage future quantum hardware to support larger molecular datasets like ChEMBL.

References

1. J. Li, R. Topaloglu, and S. Ghosh, "Quantum Generative Models for Small Molecule Drug Discovery," *arXiv preprint* arXiv:2101.03438, 2021. https://arxiv.org/pdf/2101.03438
2. N. De Cao and T. Kipf, "MolGAN: An Implicit Generative Model for Small Molecular Graphs," *arXiv preprint* arXiv:1805.11973, 2018. https://arxiv.org/pdf/1805.11973
3. S. Lloyd and C. Weedbrook, "Quantum Generative Adversarial Learning," *Physical Review Letters,* vol. 121, no. 4, p. 040502, 2018. https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.121.040502

Acknowledgements

Collaborators and Advisors:

- Special thanks to Fan Chen for their guidance and feedback during this research.

Resources:

- Used a personal NVIDIA 4090 GPU for training the models for a total time of three days.

Contact Information:

- Wali Mirza, wmirza@iu.edu