Outline of Paper: Drug Molecule Generation with VAE

Abstract

This project implements a Variational Autoencoder (VAE) for generating drug molecules. The VAE model is designed to encode molecular structures into a latent space and decode them back into molecular structures. Additionally, the model predicts properties of the generated molecules.

1 Introduction

Generative models, particularly VAEs, have become increasingly important in drug discovery for generating novel molecules, with significant contributions from SMILES-based and graph-based approaches.

2 Methodology

2.1 Data Preparation
- Data Loading and Processing
- Vocabulary Construction and Encoding

2.2 Model Architecture
- VAE Overview
- Encoder
- Decoder
- Property Predictor

2.3 Loss Function
- Reconstruction Loss
- KL Divergence
- Property Prediction Loss

3 Training Procedure

3.1 Training Setup
3.2 Training Process

4 Latent Space Optimization

4.1 Objective Function
4.2 Optimization Strategy
4.3 Decoding Process

5 Results

- Training Performance
- Latent Space Structure
- Molecule Generation

6 Discussion

- Effectiveness of the VAE
- Challenges and Limitations
- Potential Improvements

7 Conclusion

- Summary

- Future Directions

References

Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., SánchezLengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T. D., Adams, R. P., & Aspuru-Guzik, A. (2018). Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Central Science*, 4(2), 268-276. DOI: 10.1021/acscentsci.7b00572

De Cao, N., & Kipf, T. (2018). MolGAN: An implicit generative model for small molecular graphs. *arXiv preprint arXiv:1805.11973*.