Graph Neural Networks for Drug Discovery: A Comprehensive Survey

Abstract

Graph neural networks (GNNs) have emerged as a powerful tool for drug discovery, particularly in molecular property prediction and drug-target interaction modeling. This paper presents a comprehensive survey of GNN applications in pharmaceutical research, focusing on methodologies, datasets, and performance benchmarks.

Introduction

The pharmaceutical industry faces significant challenges in drug discovery, with traditional methods requiring extensive time and resources. Machine learning approaches, particularly graph neural networks, offer promising solutions by leveraging molecular graph representations to predict drug properties and interactions. Graph neural networks excel at capturing complex molecular structures by treating atoms as nodes and bonds as edges. This representation allows for the modeling of intricate chemical relationships that are crucial for understanding drug behavior.

Methodologies

2.1 Graph Convolution Networks

Graph Convolution Networks (GCNs) represent the foundation of most molecular property prediction models. These networks aggregate information from neighboring atoms to create comprehensive molecular representations.

2.2 Graph Attention Networks

Graph Attention Networks (GATs) introduce attention mechanisms to weight the importance of different molecular fragments. This approach has shown superior performance in drug-target interaction prediction tasks.

2.3 Message Passing Networks

Message Passing Neural Networks (MPNNs) provide a general framework for molecular property prediction by enabling flexible information exchange between atoms.

Datasets and Benchmarks

Several benchmark datasets are commonly used for evaluating GNN performance: • QM9: Contains 130,000 small organic molecules with quantum mechanical properties • ZINC: Large database of commercially available compounds • ChEMBL: Bioactivity database with over 2 million compounds • DrugBank: Comprehensive resource containing drug and target information

3.2 Performance Benchmarks

Recent studies have demonstrated that GNN-based approaches consistently outperform traditional molecular descriptors: • Molecular property prediction: R^2 values ranging from 0.85 to 0.95 • Drug-target interaction: AUC scores above 0.90 • ADMET prediction: Accuracy improvements of 10-15% over baseline methods

Results and Analysis

Our analysis reveals that attention-based GNNs achieve the best performance across most drug discovery tasks. The integration of molecular fingerprints with graph representations further enhances predictive accuracy. Modern GNN architectures demonstrate excellent scalability, processing large molecular databases within reasonable computational budgets.

Conclusion

Graph neural networks have revolutionized computational drug discovery by providing powerful tools for molecular representation learning. The combination of sophisticated architectures, comprehensive datasets, and rigorous benchmarking has established GNNs as the state-of-the-art approach for pharmaceutical applications.