Molecular Representation Methods in Graph Neural Networks: A Comprehensive Review

Abstract:

This review provides a systematic analysis of molecular representation methods used in Graph Neural

Introduction:
Molecular representation is fundamental to the success of machine learning approaches in drug disco

Representation Methods:

1. SMILES Strings:
- Linear text representation
- Widely used in cheminformatics
- Limited structural information
- Sequence-based processing

2. Molecular Fingerprints:
- Binary feature vectors
- Capture structural patterns
- Fast similarity calculations
- Limited interpretability

3. Graph Representations:
- Atoms as nodes, bonds as edges
- Preserves structural information
- Natural for GNN processing
- Rich feature encoding

4. 3D Conformational Representations:
- Spatial atomic coordinates
- Captures molecular geometry
- Computationally expensive
- High information content

Advanced Methods:

1. Learned Representations:
- Autoencoder-based encoding
- Task-specific optimization
- Continuous vector spaces
- Transfer learning capabilities

2. Multi-modal Representations:
- Combining multiple encodings
- Complementary information
- Improved robustness

- Higher computational cost

Evaluation Criteria:
- Representation completeness
- Computational efficiency
- Interpretability
- Transfer learning potential

Conclusion:
Graph-based representations provide the best balance of structural information and computational effi