Graph representation learning, deep generative models on graphs, and multiresolution machine learning

Truong Son Hy

January 2022

Abstract

Graph neural networks (GNNs) utilizing various ways of generalizing the concept of convolution to graphs have been widely applied to many learning tasks, including modeling physical systems, finding molecular representations to estimate quantum chemical computation, etc. Most existing GNNs address permutation invariance by conceiving of the network as a message passing scheme, where each node sums the feature vectors coming from its neighbors. We argue that this scheme imposes a limitation on the representation power of GNNs such that each node loses their identity after being aggregated by summing. Thus, we propose a new general architecture called Covariant Compositional Networks (CCNs) in which the node features are represented by higher order tensors and transform covariantly/equivariantly according to a specific representation of the symmetry group of its receptive field. Experiments show that CCNs can outperform competing methods on standard graph learning benchmarks and on estimating the molecular properties calculated by computationally expensive Density Functional Theory (DFT). This novel machine learning approach allows scientists to efficiently extract chemical knowledge and explore the increasingly growing chemical data.

Understanding graphs in a multiscale perspective is essential for capturing the large-scale structure of molecules, proteins, genomes, etc. For this reason, we introduce Multiresolution Equivariant Graph Variational Autoencoder (MGVAE), the first hierarchical generative model to learn and generate graphs in a multiresolution and equivariant manner. MGVAE is built upon Multiresolution Graph Network (MGN), an architecture which explicitly learns a multilevel hard clustering of the vertices, leading to a true multiresolution hierarchy. MGVAE then employs the hierarchical variational autoencoder model to stochastically generate a graph in multiple resolution levels given the hierarchy of latent distributions. Our proposed framework achieves competitive results with several generative tasks including general graph generation, molecule generation, unsupervised molecular representation learning, link prediction on citation graphs, and graph-based image generation. Future applications of MGVAE range from lead optimization enhancing the most promising compounds in drug discovery to finding stable crystal structures in material science.

Multiresolution Matrix Factorization (MMF) is unusual amongst fast matrix factorization algorithms in that it does not make a low rank assumption. This makes MMF especially well suited to modeling certain types of graphs with complex multiscale or hierarchical structure. While MMF promises to yield a useful wavelet basis, finding the factorization itself is hard, and existing greedy methods tend to be brittle. Therefore, we propose a "learnable" version of MMF that carefully optimizes the factorization with a combination of Reinforcement Learning and Stiefel manifold optimization through back-propagating errors. Based on the wavelet basis produced by MMF when factorizing the normalized graph Laplacian, a wavelet network learning graphs on the spectral domain is constructed with the graph convolution defined via the sparse wavelet transform. We have shown that the wavelet basis resulted from our learnable MMF far outperforms prior MMF algorithms, and the corresponding wavelet networks yield state of the art results on standard node classification on citation graphs and molecular graph classification. This is a promising direction to understand and visualize complex hierarchical structures such as social networks and biological data.

Keywords

Graph neural networks, message passing, permutation equivariance, hierarchical generative models, matrix factorization, multiresolution analysis, graph wavelets, reinforcement learning, manifold optimization