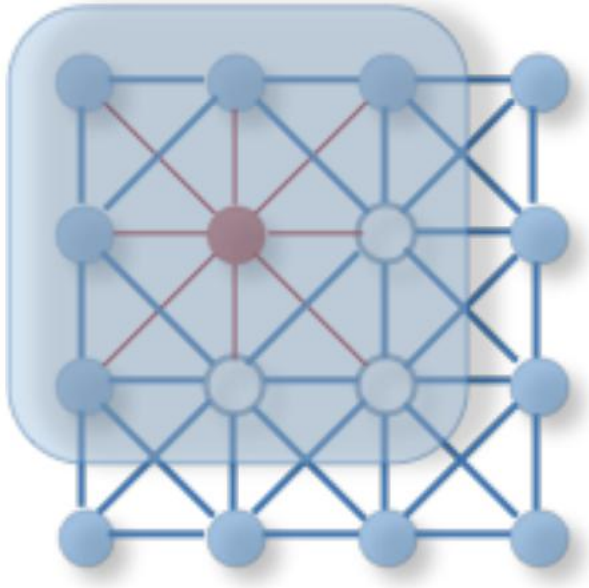
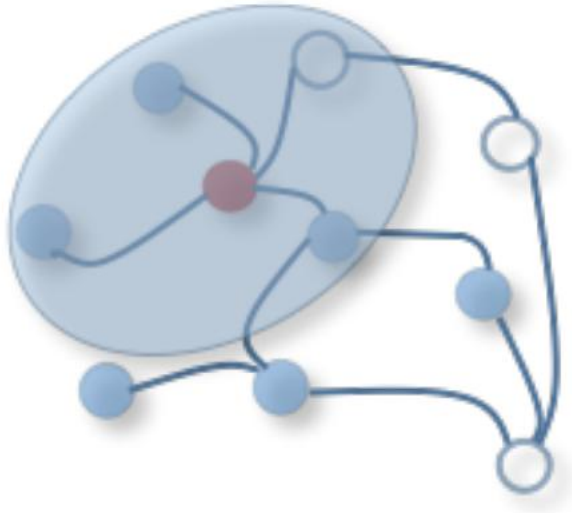
Survey on Graph Neural Networks

Motivation

Using deep learning on non-Euclidean space.





Brief History

1. Early studies learn a target node's representation by propagating neighbor information via recurrent neural architectures in an iterative manner until a stable fixed point is reached.
2. Leveraging convolutional neural network in graph domain(GCN)
3. Many other graph neural network have been proposed.

CATEGORIZATION

- Graph convolutional networks
- Graph attention networks
- Graph auto-encoders
- Graph generative networks
- Graph spatial-temporal networks

Graph Convolutional Networks

The fundamental of many complex graph neural network models.

1. Definition on spectral-domain

Introducing filters from the perspective of graph signal processing where the graph convolution operation is interpreted as removing noise from graph signals.

a. Normalized Graph Laplacian Matrix

$$L = I_n - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$$

real, symmetric, positive, semidefinite

b. Factoring Laplacian Matrix

$$L = U\Lambda U^T$$

$$U = [u_0, u_1, \dots, u_{n-1}] \in R^{n \times n}$$

Thus, vectors in U form a orthonormal space. In graph signal processing, a graph signal $x \in R^N$ is a feature vector of a node.

c. Graph Fourier Transformer

For signal x, graph Fourier transformer is defined as $\mathcal{F}(x) = U^T x$ and the inverse graph Fourier transformer is defined as $\mathcal{F}^{-1}(x) = U\hat{x}$,where $\hat{x}$ denotes the resulting signal after graph Fourier transformer .Type equation here.

d. Graph Convolutional Networks

Input: x

Filter: g

$$x *_G g = \mathcal{F}^{-1}\big(\mathcal{F}(x) \odot \mathcal{F}(g)\big) = U(U^T x \odot U^T g)$$

If denote $g_\theta = diag(U^T g)$

$$x *_G g_\theta = U_{g_\theta} U^T x$$

Key differences of Spectral-based GCN is the choice of the filter.

Summary:

Spectral CNN relies on the eigen-decomposition of the Laplacian matrix. It has three effects.

1. Any perturbation to a graph results in a change of eigenbasis.
2. The learned filters are domain dependent, meaning they cannot be applied to a graph with a different structure.
3. Eigen-decomposition requires $o(n^3)$computation and $o(n^2)$memory.

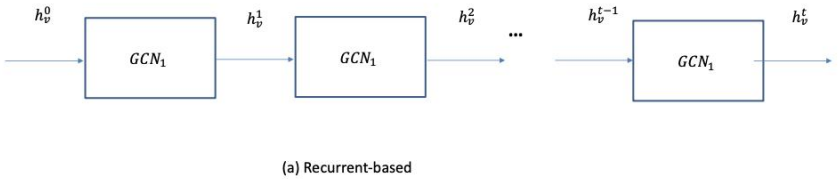
2. Definition on spatial-domain

Spatial-based approaches formulate graph convolutions as aggregating feature information from neighbors.

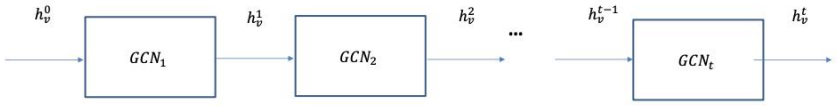
Mainly divided into two categories：

Recurrent-based :applying the same graph convolutional layer to update hidden representations.

Composition-based: applying different graph convolutional layer to update hidden representations.



(a) Recurrent-based



(b) Composition-based

a. Recurrent Based Spatial GCNs

The main idea of recurrent-based methods is to update a node's latent representation recursively until a stable fixed point is reached.

ALGORITHM 1: Learning with Stochastic Fixed Point Iteration [20]

Initialize parameters, $\{\mathbf{h}_v^0\}_{v \in \mathbf{V}}$
for $k = 1$ ***to*** K **do**
  for $t = 1$ ***to*** T **do**
    Sample n nodes from the whole node set $\mathbf{V}$
    Use Equation 10 to update hidden representations of sampled n nodes
  end
  for $p = 1$ ***to*** P **do**
    Sample m nodes from the labeled node set $\mathbf{V}$
    Forward model according to Equation 10
    Back-propagate gradients
  end
end

b. Composition Based Spatial GCNs

Composition-based methods update the nodes' representations by stacking multiple graph convolution layers.

$$\mathbf{h}_v^t = \sigma(\mathbf{W}^t \cdot aggregate_t(\mathbf{h}_v^{t-1}, \{\mathbf{h}_u^{t-1}, \forall u \in \mathcal{N}(v)\})$$

3. Graph Pooling Modules

Graph pooling module could easily reduce the variance and computation complexity by down-sampling from original feature data. Mean/max/sum pooling is the most primitive and most effective way of implementing this since calculating the mean/max/sum value in the pooling window is rapid.

4. Comparison Between Spectral and Spatial Models

Insufficient of spactral-based meth:

a. Efficiency

b. Generality

c. Flexibility

Graph Attention Networks

GAT:

A spatial-based graph convolution network where the attention mechanism is involved in determining the weights of a node's neighbors when aggregating feature information.

$$\mathbf{h}_i^t = \sigma(\sum_{j \in \mathcal{N}_i} \alpha(\mathbf{h}_i^{t-1}, \mathbf{h}_j^{t-1}) \mathbf{W}^{t-1} \mathbf{h}_j^{t-1})$$

Multi-head version:

$$\mathbf{h}_i^t = \|_{k=1}^{K} \sigma(\sum_{j \in \mathcal{N}_i} \alpha_k(\mathbf{h}_i^{t-1}, \mathbf{h}_j^{t-1}) W_k^{t-1} \mathbf{h}_j^{t-1})$$

Summary

- Attention mechanisms contribute to graph neural networks in three different ways, namely assigning attention weights to different neighbors when aggregating feature information, ensembling multiple models according to attention weights, and using attention weights to guide random walks.

Graph Auto-encoder

Graph auto-encoders are on class of network embedding approaches, Decoder reconstructs a node's neighborhood statistics such as positive pointwise mutual information (PPMI) or the first and second order of proximities.

GAE:

Encoder:

$$\mathbf{Z} = GCN(\mathbf{X}, \mathbf{A})$$

Decoder:

$$\hat{\mathbf{A}} = \sigma(\mathbf{Z}\mathbf{Z}^{\mathbf{T}})$$

- The GAE can be trained in a variational manner, i.e., to minimize the variational lower bound L:

$$L = E_{q(\mathbf{Z}|\mathbf{X},\mathbf{A})}[\log_p(\mathbf{A}|\mathbf{Z})] - KL[q(\mathbf{Z}|\mathbf{X},\mathbf{A})||p(\mathbf{Z})]$$

- DNGR

Graph Generative Networks

- The goal of graph generative networks is to generate graphs given an observed set of graphs.
- Molecular Generative Adversarial Networks (MolGAN)
- Deep Generative Models of Graphs (DGMG)
- GraphRNN

Graph Spatial-Temporal Networks

- Graph spatial-temporal networks capture spatial and temporal dependencies of a spatial-temporal graph simultaneously. Spatial-temporal graphs have a global graph structure with inputs to each node which are changing across time. The goal of graph spatial-temporal networks can be forecasting future node values or labels, or predicting spatial-temporal graph labels.

DATASETS

- Citation Networks: Cora, Citeseer and Pubmed
- Social Networks: BlogCatalog, Reddit, and Epinions.
- Chemical/Biological Graphs: NCI-1 , NCI-9, MUTAG, D&D, QM9
- Others: METR-LA, MovieLens-1M, NELL

Applications

- Computer Vision
- Recommender Systems
- Traffic
- Chemistry

FUTURE DIRECTIONS

- Go Deep
- Receptive Field
- Scalability
- Dynamics and Heterogeneity