Dwivedi et al. 2020: Benchmarking Graph Neural Networks

Minqi Pan

April 10, 2020

Graph Neural Ordinary Differential Equations

- arXiv:2003.00982, Mar 2nd, 2020
- Vijay Prakash Dwivedi, Chaitanya K. Joshi, Thomas Laurent, Yoshua Bengio, Xavier Bresson
- Nanyang Technological University, Loyola Marymount University, Montreal Institute for Learning Algorithms, Université de Montréal, CIFAR

Outline

1. Background
   - Isotropic, Anisotropic and Hierarchical Models
2. Proposals
   - Proposed Benchmarking Framework
3. Discoveries
   - Issues with CORA and TU Datasets
   - Numerical Experiments
   - What did we learn?

Outline

"Isotropic" Models

Each neighbor contributes equally to the update of the central node, treating every "edge direction" equally.

GCN: Iteratively Updating Node Representations

$$\hat{h}_i^{\ell+1} = \frac{1}{\text{deg}_i} \sum_{j \in \mathcal{N}_i} h_j^{\ell}, \quad h_i^{\ell+1} = \sigma(U^{\ell} \hat{h}_i^{\ell+1})$$

- Sukhbaatar et al. 2016, Kif and Welling 2017
- h_i^{ℓ}: the d-dimensional embedding representation of node i at layer $\ell + 1$
- $\mathcal{N}_i$: the set of nodes connected to node i on the graph
- $\text{deg}_i = |\mathcal{N}_i|$: the degree of node i
- σ: a nonlinearity
- $U^{\ell} \in \mathbb{R}^{d \times d}$: a learnable parameter

GraphSage: Variations of Averaging Mechanism

$$\hat{h}_i^{\ell+1} = \text{Concat}\Big(h_i^\ell, \frac{1}{\text{deg}_i} \sum_{j \in \mathcal{N}_i} h_j^\ell\Big)$$

- Hamilton et al. 2017
- The embeddings vectors are projected onto the unit ball before being passed to the next layer

Graph-Isomorphism-Network: Another Variation

$$\hat{h}_i^{\ell+1} = (1+\epsilon)h_i^\ell + \sum_{j \in \mathcal{N}_i} h_j^\ell,$$
$$h_i^{\ell+1} = \sigma\big(U^\ell \sigma(\mathsf{BN}(V^\ell \hat{h}_i^\ell + 1))\big)$$

- Xu et al. 2019
- ϵ, U^ℓ, V^ℓ: learnable parameters
- BN: Batch Normalization
- GIN uses the features at all intermediate layers for the final prediction

"Anisotropic" Models

$$\hat{h}_i^{\ell+1} = w_i^\ell h_i^\ell + \sum_{j \in \mathcal{N}_i} w_{ij}^\ell h_j^\ell$$

- w_i^ℓ, w_{ij}^ℓ: weights that are computed using attention or gating mechanisms
- MoNet: Gaussian Mixture Model Networks (Monti et al. 2017)
- GatedGCN: Graph Convolutional Networks (Bresson & Laurent, 2017)
- GAT: Graph Attention Networks (Veličković et al. 2018)

“Hierarchical” Models

- Ying et al. 2018
- DiffPool: Differentiable Pooling, using the GraphSage formulation at each stage of the hierarchy and for the pooling

Outline

Proposed Benchmark Datasets

- Computer Vision: graphs constructed with super-pixels
  1. MNIST: 70k graphs w/ 40-75 nodes
  2. CIFAR10: 60k graphs w/ 85-150 nodes
- Chemistry: real-world molecular graphs
  1. ZINC: 12k graphs w/ 9-37 nodes
- Artificial: graphs generated from stochastic block model or from uniform distribution
  1. PATTERN: 14k graphs w/ 50-180 nodes
  2. CLUSTER: 12k graphs w/ 40-190 nodes
  3. TSP: 12k graphs w/ 50-500 nodes

Outline

1. Background
   - Isotropic, Anisotropic and Hierarchical Models
2. Proposals
   - Proposed Benchmarking Framework
3. Discoveries
   - Issues with CORA and TU Datasets
   - Numerical Experiments
   - What did we learn?

Outline

Outline