Symmetry kernels for graph classification

Yannick Kuhar

July 11, 2024

Introduction

- data naturally appears as graphs in many domains
- the size of the graphs can pose computational issues
- for some algorithms to work efficiently, we must reduce them

Graph reduction techniques

- Graph Compression
- combining nodes, also known as Graph Coarsening
- removing redundant edges, also known as Graph Sparsification

Graph reduction techniques

- 4 methods
- Graphlet-based compression
- Graph reduction with spectral and cut guarantees
- Principle of Relevant Information (PRI) for Graph Sparsification
- Graph Coarsening with Machine Learning

GRAPHLET-BASED COMPRESSION

- list of small graph, i.e., graphlets that exhibit symmetry compressibility
- algorithm identifies the maximal set of subgraphs isomorphic to each graphlet
- all such subgraphs are removed
- only the non-redundant parts of the graphs are kept

Graph reduction with spectral and cut guarantees

- graph Laplacian: $L = D - A$
- D is the diagonal matrix of node degrees, and A is the adjacency matrix
- reduction scheme: $L_l = P_l^{\mp} L_{l-1} P_l^{+}$ and $x_l = P_l x_{l-1}$
- $L_0 = L$, P is the reduction matrix and $\mp$ denotes the pseudoinverse
- this reduces the graph while keeping all its properties

PRI for Graph Sparsification

- traditionally, this was done by using the Information bottleneck formula
- Yu et al. explored the use of the principle of relevant information
- this reduces the number of random variables
- the resulting equations are solved by a gradient descent algorithm

Graph Coarsening with Machine Learning

- node merging candidate sets determined by KNN
- this is done on an attribute matrix of the nodes, not the graph structure
- algorithm then calculates the feature vector for the super nodes

Results

Algorithm	$\|V\|$	$\|E\|$	Reduction rate $\|V\|$	Reduction rate $\|E\|$
Original	25	39	0.00%	0.00%
Cytocoarsening	30	88	20.00%	125.64%
Graph-coarsening-1.1	10	13	-60.00%	-66.67%
Symmetry compression	17	12	-32.00%	-69.23%
GraphPRI-0-0	19	17	-24.00%	-56.41%
GraphPRI-0-2	16	27	-36.00%	-30.77%
GraphPRI-0-3	17	29	-32.00%	-25.64%
GraphPRI-0-4	14	25	-44.00%	-35.90%
GraphPRI-0-5	30	66	20.00%	69.23%
GraphPRI-0-1000	32	73	28.00%	87.18%
GraphPRI-0.001-2	27	60	8.00%	53.85%
GraphPRI-0.001-1000	32	74	28.00%	89.74%

Conclusions

- graphlet compression has reduced the graph to three components
- coarsening with ML inflated the graph due to a lack of attribute matrix
- graph reduction with spectral and cut guarantees yields a reduction that could be useful
- GraphPRI yielded the two most promising reductions
- test neural network-based reductions