Rethinking the Capacity of Graph Neural Networks for Branching Strategy

Ziang Chen[1]

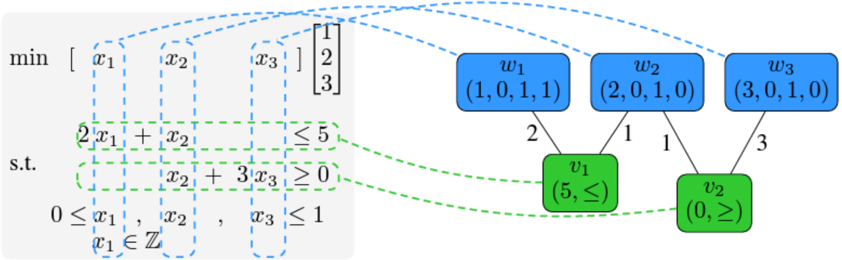
Joint work with Jialin Liu, Xiaohan Chen, Xinshang Wang, and Wotao Yin

[1]Department of Mathematics, Massachusetts Institute of Technology

38th Conference on Neural Information Processing Systems
Vancouver, Canada

Mixed-Integer Linear Programs and Graph Representation

- Mixed-Integer Linear Program (MILP): optimizing linear objective function subject to linear and integer constraints.
- The information in a MILP problem can be encoded into a weighted bipartite graph with vertex features (Gasse et al., 2019):



Strong Branching

- Strong branching score $\mathrm{SB}(G) \in \mathbb{R}^n$.
  - A widely used heuristic that effectively reduces the size of the branch-and-bound (BnB) search space.
  - Computationally expensive (solving $\mathcal{O}(n)$ linear programs (LPs)).
- $x^*_{\mathrm{LP}}(G) \in \mathbb{R}^n$ is the optimal solution with the smallest ℓ_2-norm to the LP relaxation.
- If x_j is not an integer variable, then $\mathrm{SB}(G)_j = 0$.
- If x_j is an integer variable, then

$$\mathrm{SB}(G)_j = (f^*_{\mathrm{LP}}(G, j, l_j, \hat{u}_j) - f^*_{\mathrm{LP}}(G)) \cdot (f^*_{\mathrm{LP}}(G, j, \hat{l}_j, u_j) - f^*_{\mathrm{LP}}(G)),$$

where $\hat{u}_j = \lfloor x^*_{\mathrm{LP}}(G)_j \rfloor$, $\hat{l}_j = \lceil x^*_{\mathrm{LP}}(G)_j \rceil$, and f^*_{LP} is the optimal objective value of the LP relaxation.

Message-Passing Graph Neural Networks

- Message-passing graph neural networks (MP-GNNs)
- Message-passing layers

$$s_i^\ell = p^\ell\left(s_i^{\ell-1}, \sum_{j=1}^{n} E_{i,j} f^\ell(t_j^{\ell-1})\right), \quad t_j^\ell = q^\ell\left(t_j^{\ell-1}, \sum_{i=1}^{m} E_{i,j} g^\ell(s_i^{\ell-1})\right).$$

- For any $\epsilon, \delta > 0$ and any MILP data distribution $\mathbb{P}$ supported on "MP-tractable" instances, there exists an MP-GNN F such that

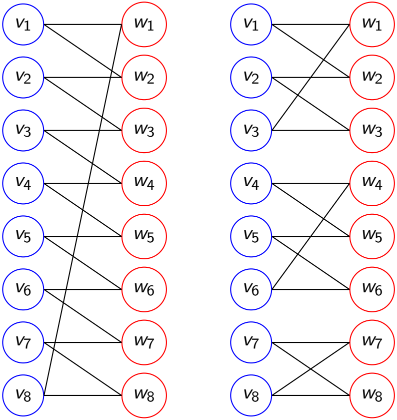
$$\mathbb{P}[\|F(G) - \mathrm{SB}(G)\| \le \delta] \ge 1 - \epsilon.$$

- MP-tractability: edges with the same pair of vertex features have the same weight.
- A generic MILP instance is MP-tractable.
- For non-MP-tractable MILPs, MP-GNNs may fail to represent SB.

A Counter-Example for MP-GNNs

$$\begin{aligned}
\min \quad & x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8, \\
\text{s.t.} \quad & x_1 + x_2 \geq 1,\ x_2 + x_3 \geq 1,\ x_3 + x_4 \geq 1, \\
& x_4 + x_5 \geq 1,\ x_5 + x_6 \geq 1,\ x_6 + x_7 \geq 1, \\
& x_7 + x_8 \geq 1,\ x_8 + x_1 \geq 1, \\
& 0 \leq x_j \leq 1,\ x_j \in \mathbb{Z},\ 1 \leq j \leq 8
\end{aligned}$$

$$\begin{aligned}
\min \quad & x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8, \\
\text{s.t.} \quad & x_1 + x_2 \geq 1,\ x_2 + x_3 \geq 1,\ x_3 + x_1 \geq 1, \\
& x_4 + x_5 \geq 1,\ x_5 + x_6 \geq 1,\ x_6 + x_4 \geq 1, \\
& x_7 + x_8 \geq 1,\ x_8 + x_7 \geq 1, \\
& 0 \leq x_j \leq 1,\ x_j \in \mathbb{Z},\ 1 \leq j \leq 8.
\end{aligned}$$

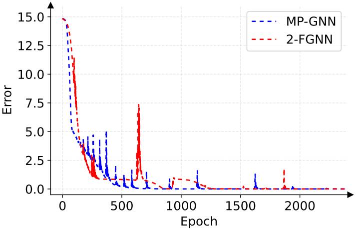


Second-Order Folklore Graph Neural Networks

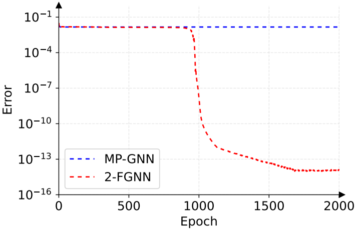
- Second-Order Folklore Graph Neural Networks (2-FGNNs)
- Computation via edge features.
- Internal layers:
  - $s_{ij}^{l} = p^{l}\big(s_{ij}^{l-1}, \sum_{j_1 \in W} f^{l}(t_{j_1 j}^{l-1}, s_{ij_1}^{l-1})\big)$ for all $i \in V, j \in W$, and
  - $t_{j_1 j_2}^{l} = q^{l}\big(t_{j_1 j_2}^{l-1}, \sum_{i \in V} g^{l}(s_{ij_2}^{l-1}, s_{ij_1}^{l-1})\big)$ for all $j_1, j_2 \in W$.
- Final layer:
  - $y_j = r\big(\sum_{i \in V} s_{ij}^{L}, \sum_{j_1 \in W} t_{j_1 j}^{L}\big)$.
- For any $\epsilon, \delta > 0$ and any MILP data distribution $\mathbb{P}$, there exists an 2-FGNN F such that

$$\mathbb{P}[\|F(G) - \mathrm{SB}(G)\| \leq \delta] \geq 1 - \epsilon.$$

Numerical Results



(a) Random dataset



(b) Dataset with symmetry

The End

Thanks for your listening!