Region and Relations Based Multi Attention Network for Graph Classification

Manasvi Aggarwal, M. N. Murty

Myntra & Indian Institute of Science, Bangalore - 560034

Motivations behind this work

- Most of the existing graph pooling techniques for graph classification can't handle long-range dependencies between nodes.
- On the other hand, the node's co-relations with other nodes are also important for a more expressive model.
- Further, the existing pooling approaches are either global (cannot preserve the structure) or hierarchical (maintain the hierarchical structure).
- Hierarchical pooling techniques use a single classifier at the end of hierarchical structure which is a cause of information squandering.

Our Contributions

We propose a multi-attention network **R2MAN** which:

- includes our proposed **R2POOL** layer that forms the coarser version of the graph based on our proposed **region based attention** and **relation aware attention** layers.
- combines R2POOL layer with our **attention - aware multi-level prediction** mechanism to learn coarse to fine representations and restrict them to use only intermediate features weighted by the **alignment scores**.
- leverages the proposed **branch training** to learn importance of each level prediction.

Hierarchical training strategy

- Compatibility scores (C^p) between each pth intermediate layer representations (X^p) and vector of final level graph (O^P):

$$e_i^p = x_i^p O^P, \quad c_i^p = \frac{exp(e_i^p)}{\sum_{j=1}^{N_p} exp(e_j^p)}$$

$$l^p = C^p X^p \quad or \quad l_p = \sum_{i=1}^{N_p} c_i^p x_i^p$$

- Branch nets at each intermediate layer:

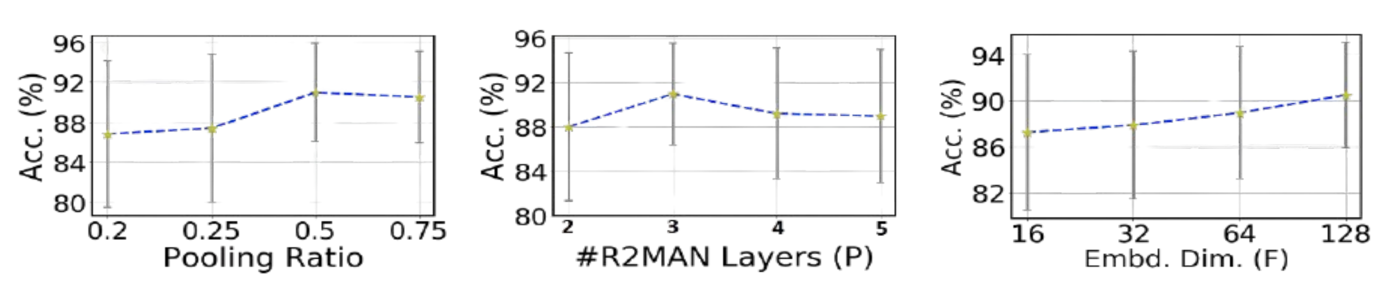
$$S_p = Branch\_net(l_p)$$

- Finally, branch training is used to get the final predictions (S_{model}):

$$S_{model} = \sum_{p=1}^{P-1} V_p S_p$$

Sensitivity Analysis

Sensitivity analysis of R2MAN with respect to various hyper-parameters . All the variations are as expected which shows the robustness of R2MAN.

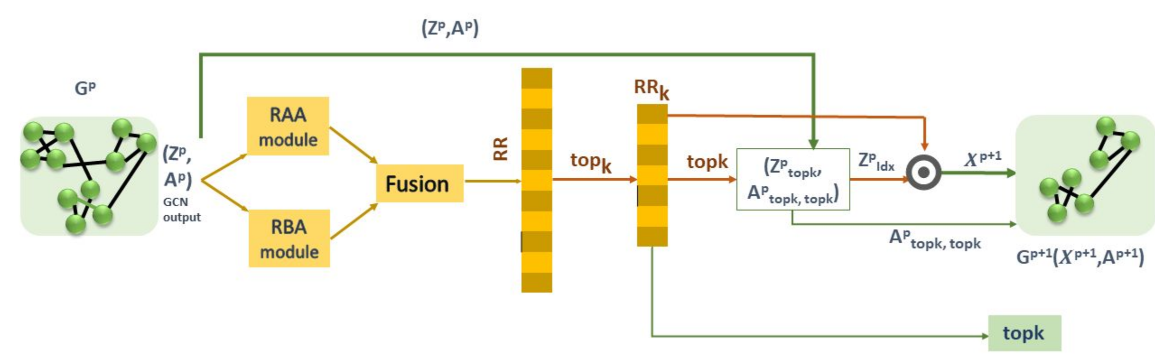


(a) (b) (c)

Conclusion

Through various experiments, we show that R2MAN learns improved graph level embeddings with improved performance on many graph datasets.

Architecture of R2POOL Layer



- RBA (Region Based Attention) which calculates importance of all nodes (LA) considering only their local region uses a single graph neural network layer.
- RAA (Relation Aware Attention) calculates importance of nodes at graph level (RA) using self attention operator:

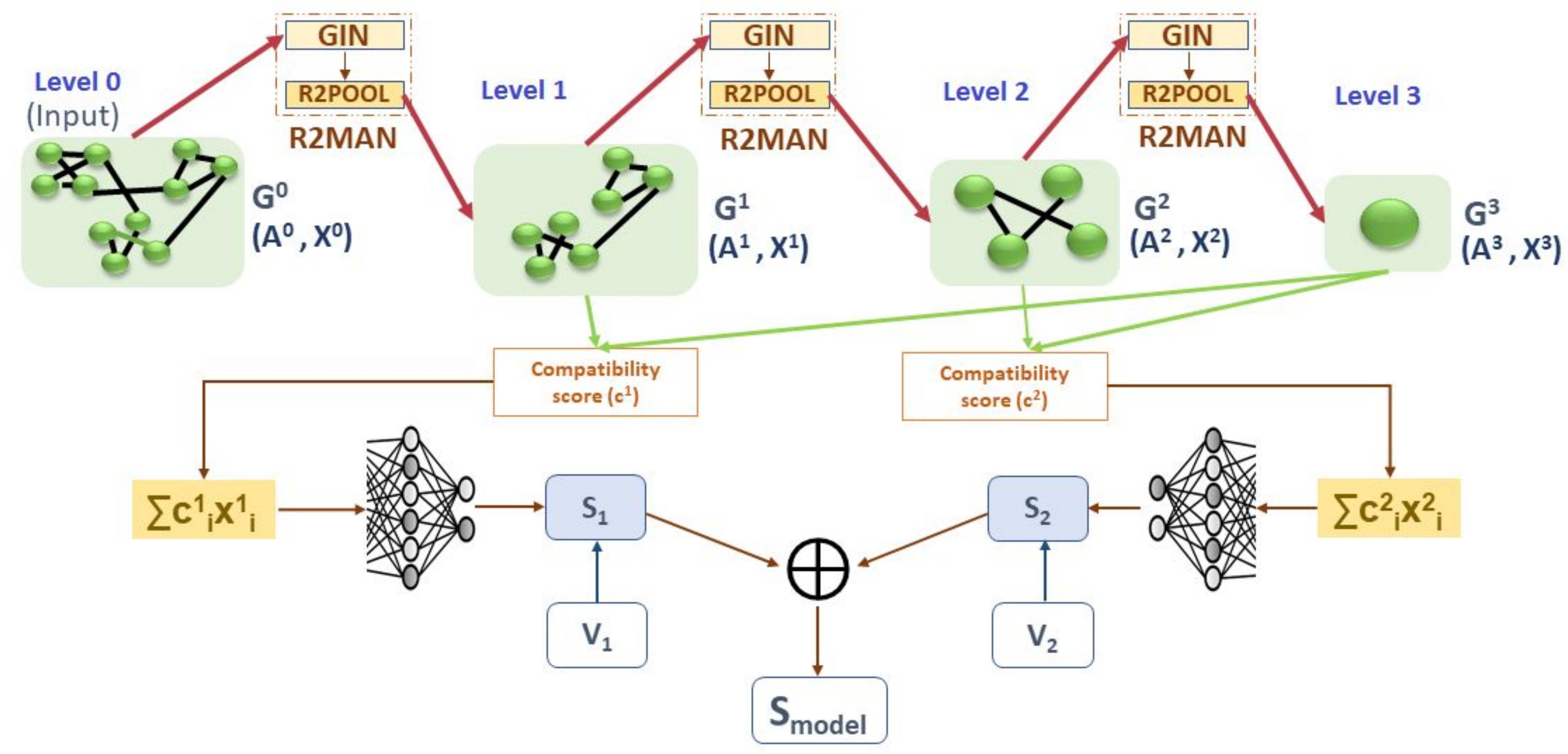
$$R_{ij} = f(z_i^p)^T g(z_j^p)$$

$$r_{ij} = \frac{exp(R_{ij})}{\sum_{q=1}^{N_p} exp(R_{iq})}, \; RS_i = \sum_{q=1}^{N_p} r_{qi} \qquad c_{ij} = \frac{exp(R_{ij})}{\sum_{q=1}^{N_p} exp(R_{qi})}, \; CS_i = \sum_{q=1}^{N_p} c_{iq} \qquad \forall i \in \{1, \cdots, N_p\}$$

Final vector of scores is: **RA = σ (RS+CS)**.

- Combines these attention scores to get final scores.: **RR=σ(α(LA+RA)+β(LA×RA)).**
- Finally, selects k nodes to form a coarser version of the graph using Top-k Sampling.

Complete Architecture of R2MAN



Graph Classification Results (Acc. and Std. Dev.)

Algorithms	MUTAG	PTC	PROTEINS	NCI1	NCI109	IMDB-B	IMDB-M
GK	81.39±1.7	55.65±0.5	71.39±0.3	62.49±0.3	62.35±0.3	NA	NA
RW	79.17±2.1	55.91±0.3	59.57±0.1	NA	NA	NA	NA
PK	76±2.7	59.5±2.4	73.68±0.7	82.54±0.5	NA	NA	NA
WL	84.11±1.9	57.97±2.5	74.68±0.5	**84.46±0.5**	**85.12±0.3**	NA	NA
AWE-DD	NA	NA	NA	NA	NA	74.45±5.8	51.54±3.6
AWE-FB	87.87±9.7	NA	NA	NA	NA	73.13±3.2	51.58±4.6
node2vec	72.63±10.20	58.85±8.00	57.49±3.57	54.89±1.61	52.68±1.56	NA	NA
sub2vec	61.05±15.79	59.99±6.38	53.03±5.55	52.84±1.47	50.67±1.50	55.26±1.54	36.67±0.83
graph2vec	83.15±9.25	60.17±6.86	73.30±2.05	73.22±1.81	74.26±1.47	71.1±0.54	50.44±0.87
InfoGraph	89.01±1.13	61.65±1.43	NA	NA	NA	73.03±0.87	49.69±0.53
SortPool	85.83±1.7	58.59±2.5	75.54±0.9	74.44±0.5	72.31	70.03±0.9	47.83±0.9
PSCN	88.95±4.4	62.29±5.7	75±2.5	76.34±1.7	NA	71±2.3	45.23±2.8
DCNN	NA	NA	61.29±1.6	56.61±1.0	NA	49.06±1.4	33.49±1.4
ECC	76.11	NA	NA	76.82	75.03	NA	NA
DGK	87.44±2.7	60.08±2.6	75.68±0.5	80.31±0.5	80.32±0.3	66.96±0.6	44.55±0.5
DIFFPOOL	85.56	62.8	76.25	NA	NA	74.3	50.3
SAGPool	81.9	61.6	72.1	74.2	74.1	72.2	50.4
gpool	80.3	NA	77.7	NA	NA	73.0	49.9
IGN	83.89±12.95	58.53±6.86	76.58±5.49	74.33±2.71	72.82±1.45	72.0±5.54	48.73±3.41
GIN	89.4±5.6	64.6±7.0	76.2±2.8	82.7±1.7	NA	75.1±5.1	**52.3±2.8**
1-2-3GNN	86.1±	60.9±	75.5±	76.2±	NA	74.2±	49.5±
R2MAN	**92.11±5.35**	**64.90±5.99**	**77.84±1.51**	79.01±2.53	77.80±1.73	**75.83±3.17**	51.80±3.35
Rank	1	1	1	5	3	1	2

t-SNE visualization (MUTAG dataset)

The best performance is achieved when embeddings are generated by complete model, R2MAN. Moreover, R2MAN-I and R2MAN-B are still performing better than other 2 baselines.

(a) DIFFPOOL (b) SAGPool (c) R2MAN-I (d) R2MAN-B (e) R2MAN

Future Direction

In future, we would like to explore new methods for relations aware attention scores of nodes i.e., the RAA module.

Source Code and Contact

Source Code of R2MAN & datasets: https://github.com/manasviaggarwal/R2MAN. Contact: manasvi.aggarwal@myntra.com