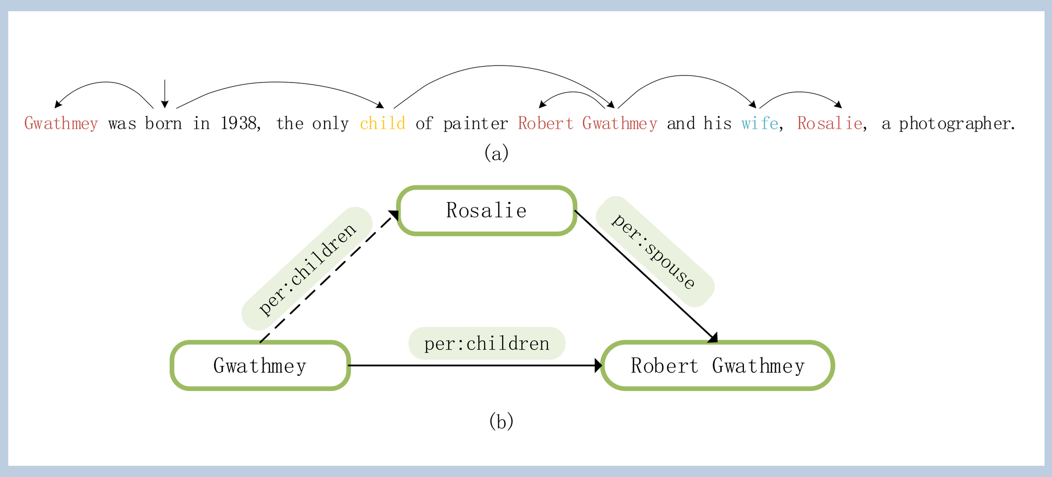
Adaptive Graph Convolutional Networks with Attention Mechanism for Relation Extraction

Zhixin Li, Yaru Sun, Suqin Tang, Canlong Zhang, Huifang Ma
lizx@gxnu.edu.cn, MiniiEcho@163.com, sqtang@gxnu.edu.cn, clzhang@gxnu.edu.cn, mahuifang@yeah.net

Problem

The existing models adopt an end-to-end approach can be divided into two categories: the sequence-based and the dependency-based model. But they have two problems.

1. rich structural information on the dependency tree.
2. multi-hop influence on the dependent path.



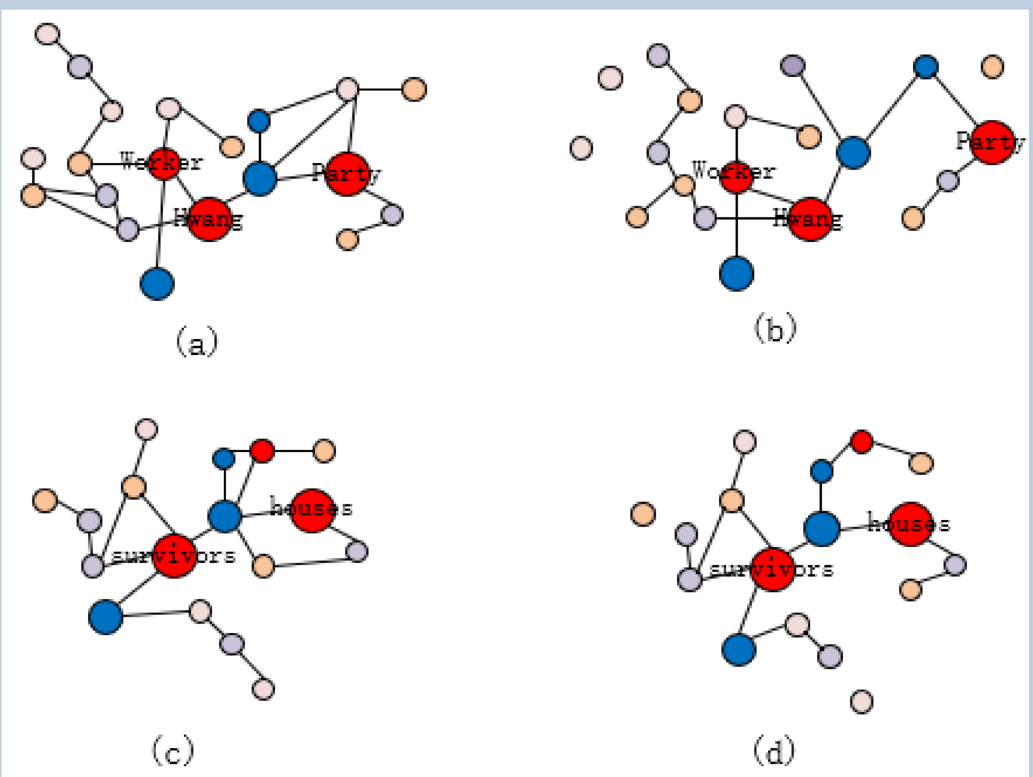
Basic Concepts

Our purpose is to get the optimal expectation value through iteration, which is obtained from the Bellaman optimization formula, $Q(h,r)$ represent the cumulative return of r when the action is executed in h state, γ is a penalty factor.

$$Q(h,r) = \mathbb{E}R(h,r) + \gamma \mathbb{E}Q(h,r)$$
$$Z(h,r) = R(h,r) + \gamma Z\left(h,r'\right)$$
$$\mathcal{L} = \sum_{s \in \mathcal{S}} \sum_{i=j} \log Z\left(h, r_{ij} \mid i,j,s\right))$$

Visualization

we visualize the nodes from the semantic association strength and classification effect. In fig (a)and(c) with the attention, (b) and (d) without. We can see that the model with the attention strengthens the semantic relevance between nodes.

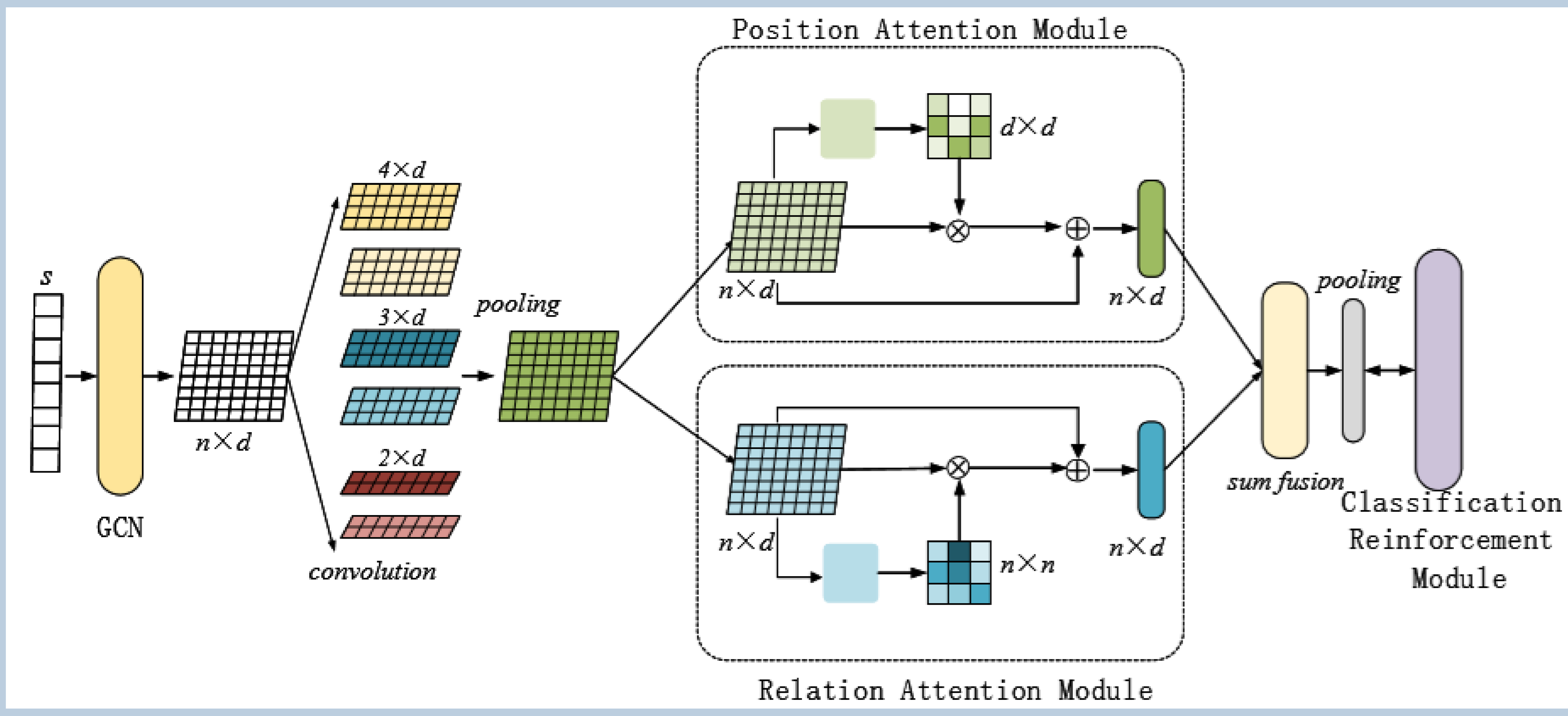


Acknowledgements

This work is supported by the National Natural Science Foundation of China (Nos. 61966004, 61663004, 61866004, 61762078, 61967002), the Guangxi Natural Science Foundation (Nos. 2019GXNSFDA245018, 2018GXNSFDA281009, 2017GXNSFAA198365), the Guangxi Bagui Scholar Teams for Innovation and Research Project, the Guangxi Talent Highland Project of Big Data Intelligence and Application, Guangxi Collaborative Innovation Center of Multi-source Information Integration and Intelligent Processing.

Dual Attention-Guided GCN (DAGCN)

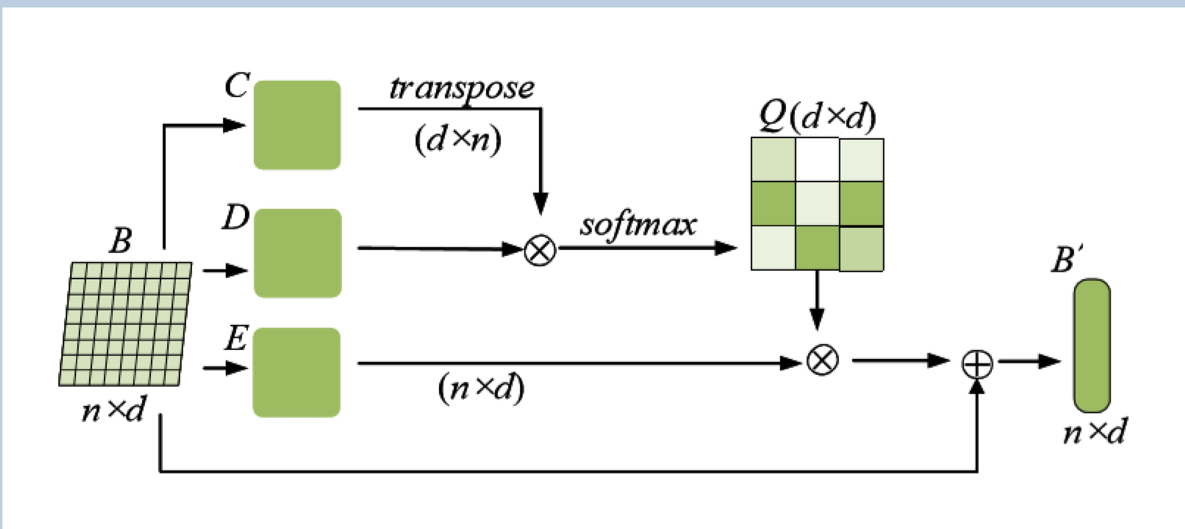
We propose a relationship extraction model that applies a dual attention mechanism with reinforcement learning in a graph convolutional network. It's called DAGCN.

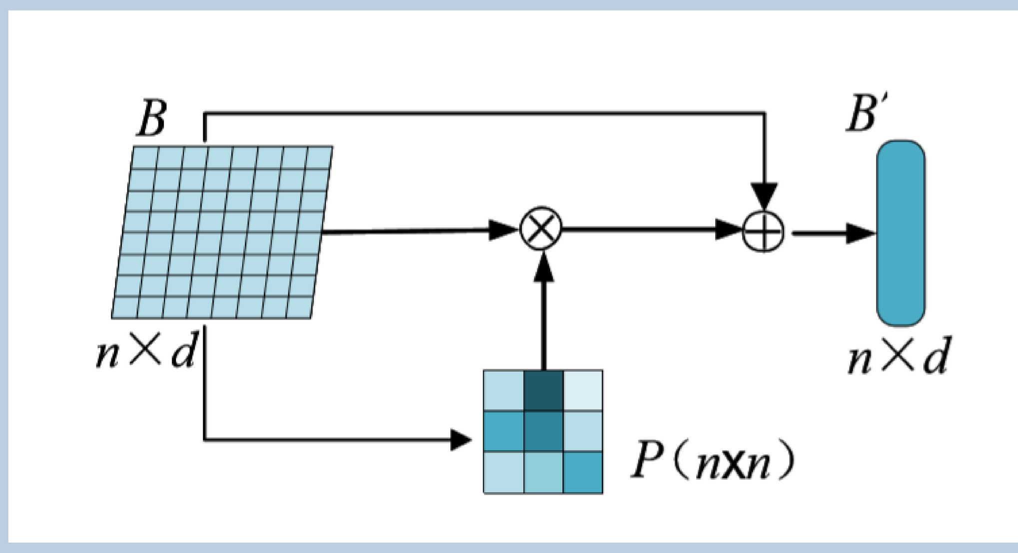


Details & Experiments

New features of spatial contextual information are generated through the following three steps. First, generate the position attention matrix, which models the spatial relationship between any two positions of the features. Then, matrix multiplication is performed between the attention matrix and the original feature. Finally, we perform an element-wise sum operation between the multiplied results and the original features to obtain the final representation of the global contextual information.

The relation attention model directly calculates the relation attention matrix P. The matrix is generated according to the dependency between nodes. Then, the relationship features are generated by self-attention mechanism. Then, we perform a matrix multiplication on attention matrix and original feature. Finally, an element-wise sum is performed between the multiplied results and the original features to obtain the global dependency between nodes.





We observed performance under varied training data sizes. Find that as the training data increase in scale, the performance gap among the three models becomes more obvious. When the size of the dataset reaches 80%, the F1 value of ours is close to the maximum value of C-AGGCN. These results show that the training dataset is more effectively utilized by the model proposed in this paper.

We observed performance under varied training sentence length. Find that as the sentence length increases, the dependency graph includes more nodes, and the information acquisition performance decreases. However, the results show that ours can better obtain more useful information with a larger dependency graph.