Learning to Execute Programs with Instruction Pointer Attention Graph Neural Networks

Challenger: Jialin Chen

2024/01/25

- Why gated graph neural network (GGNN) is selected as the base model to conduct the ablation study, instead of other message-passing neural networks? Is there any intrinsic relationship between gate mechanism and program execution?

Table 1: The IPA-GNN model is a message passing GNN. Selectively replacing its components with those of the GGNN yields two baseline models, NoControl and NoExecute. Blue expressions originate with the IPA-GNN, and orange expressions with the GGNN.

	IPA-GNN (Ours)	NoControl	NoExecute	GGNN
$h_{0,n}$	$= 0$	$= 0$	$= \mathrm{Embed}(x_n)$	$= \mathrm{Embed}(x_n)$
$a^{(1)}_{t,n}$	$= \mathrm{RNN}(h_{t-1,n}, \mathrm{Embed}(x_n))$	$= \mathrm{RNN}(h_{t-1,n}, \mathrm{Embed}(x_n))$	$= h_{t-1,n}$	$= h_{t-1,n}$
$a^{(2)}_{t,n,n'}$	$= p_{t-1,n'} \cdot b_{t,n',n} \cdot a^{(1)}_{t,n}$	$= 1 \cdot a^{(1)}_{t,n}$	$= p_{t-1,n'} \cdot b_{t,n',n} \cdot \mathrm{Dense}(a^{(1)}_{t,n})$	$= 1 \cdot \mathrm{Dense}(a^{(1)}_{t,n})$
$\tilde{h}_{t,n}$	$= \sum_{n' \in N_{\text{in}}(n)} a^{(2)}_{t,n,n'}$	$= \sum_{n' \in N_{\text{all}}(n)} a^{(2)}_{t,n,n'}$	$= \sum_{n' \in N_{\text{in}}(n)} a^{(2)}_{t,n,n'}$	$= \sum_{n' \in N_{\text{all}}(n)} a^{(2)}_{t,n,n'}$
$h_{t,n}$	$= \tilde{h}_t$	$= \tilde{h}_t$	$= \mathrm{GRU}(h_{t-1,n}, \tilde{h}_{t,n})$	$= \mathrm{GRU}(h_{t-1,n}, \tilde{h}_{t,n})$

- The paper claims that IPA-GNN has better systematic generalization to out-of-distribution programs. Which novel components in IPA-GNN potentially leads to the improved generalization?