Lattice CNNs for Matching Based Chinese Question Answering

AAAI Press
Association for the Advancement of Artificial Intelligence
2275 East Bayshore Road, Suite 160
Palo Alto, California 94303

Appendix

Formula Derivation

Lattice based CNN layer with average pooling, derived from Eq.**??**. Omit the condition $w_i \in V$, set $n = 3$ and mark $\boldsymbol{W_c}$ as $(\boldsymbol{W_{c1}} : \boldsymbol{W_{c2}} : \boldsymbol{W_{c3}})$, where the size of $\boldsymbol{W_{ci}}$ is $(m', m)\ \forall i \in \{1, 2, 3\}$, and suppose the pooling mode is average, and ignore the activation function.

$$
\begin{aligned}
F_w = \texttt{Pool}\{f(\boldsymbol{W_c}(\boldsymbol{v_{w_1}} : ... : \boldsymbol{v_{w_n}}) + \boldsymbol{b_c^T})| \\
\forall i, w_i \in V, (w_i, w_{i+1}) \in E, w_{\lceil \frac{n+1}{2} \rceil} = w\} \\
= \texttt{ave}\{(\boldsymbol{W_{c1}} : \boldsymbol{W_{c2}} : \boldsymbol{W_{c3}})(\boldsymbol{v_{w_1}} : \boldsymbol{v_{w_2}} : \boldsymbol{v_{w_3}}) + \boldsymbol{b_c^T}| \\
\forall i, (w_i, w_{i+1}) \in E, w_2 = w\} \\
= \texttt{ave}\{\boldsymbol{w_{c1}v_{w_1}} + \boldsymbol{w_{c2}v_w} + \boldsymbol{w_{c3}v_{w_3}} + \boldsymbol{b_c^T}| \\
(w_1, w), (w, w_3) \in E\} \\
= \boldsymbol{w_{c2}v_w} + \boldsymbol{b_c^T} + \frac{\sum_{w_1} \boldsymbol{w_{c1}v_{w_1}}}{t_1} + \frac{\sum_{w_3} \boldsymbol{w_{c3}v_{w_3}}}{t_2}
\end{aligned} \tag{1}
$$

, where t_1 is the number of previous words of w and t_2 is the number of next words.

The same as Eq.1, but have maximum pooling.

$$
\begin{aligned}
F_w = \texttt{Pool}\{f(\boldsymbol{W_c}(\boldsymbol{v_{w_1}} : ... : \boldsymbol{v_{w_n}}) + \boldsymbol{b_c^T})| \\
\forall i, w_i \in V, (w_i, w_{i+1}) \in E, w_{\lceil \frac{n+1}{2} \rceil} = w\} \\
= \texttt{max}\{(\boldsymbol{W_{c1}} : \boldsymbol{W_{c2}} : \boldsymbol{W_{c3}})(\boldsymbol{v_{w_1}} : \boldsymbol{v_{w_2}} : \boldsymbol{v_{w_3}}) + \boldsymbol{b_c^T}| \\
\forall i, (w_i, w_{i+1}) \in E, w_2 = w\} \\
= \texttt{max}\{\boldsymbol{w_{c1}v_{w_1}} + \boldsymbol{w_{c2}v_w} + \boldsymbol{w_{c3}v_{w_3}} + \boldsymbol{b_c^T}| \\
(w_1, w), (w, w_3) \in E\} \\
= \boldsymbol{w_{c2}v_w} + \boldsymbol{b_c^T} + \max\{\boldsymbol{w_{c1}v_{w_1}}|\forall w_1\} + \max\{\boldsymbol{w_{c3}v_{w_3}}|\forall w_3\}
\end{aligned} \tag{2}
$$

Directed graph convoluational networks (DGCs), derived from *Eq.2* in (Vashishth et al. 2018). We also suppose the pooling mode is average(not sum, which is in original paper and the first line of our derivation) and ignore the activation function.

$$
\begin{aligned}
h_v^{k+1} = f(\sum_{u \in \mathcal{N}} (\boldsymbol{W}_{l(u,v)}^k h_u^k + \boldsymbol{b}_{l(u,v)}^k)) \\
= \texttt{ave}\{(\boldsymbol{W}_{l(u,v)}^k h_u^k + \boldsymbol{b}_{l(u,v)}^k)|u \in \mathcal{N}\} \\
= \texttt{ave}\{\boldsymbol{W_{c1}V_{w_1}} + \boldsymbol{b_1^T}, \boldsymbol{W_{c3}V_{w_3}} + \boldsymbol{b_3^T}, \boldsymbol{W_{c2}V_w} + \boldsymbol{b_2^T}| \\
(w_1, w), (w, w_3) \in E\} \\
= \frac{1}{1 + t_1 + t_2}(\boldsymbol{w_{c2}v_w} + (\boldsymbol{b_2^T} + \boldsymbol{b_1^T} * \boldsymbol{t_1} + \boldsymbol{b_3^T} * \boldsymbol{t_2}) \\
+ \sum_{w_1} \boldsymbol{w_{c1}v_{w_1}} + \sum_{w_3} \boldsymbol{w_{c3}v_{w_3}})
\end{aligned} \tag{3}
$$

, where $\boldsymbol{W_{c1}}, \boldsymbol{b_1^T}$ are parameters for forward connections, $\boldsymbol{W_{c3}}, \boldsymbol{b_3^T}$ are parameters for backward connections, and $\boldsymbol{W_{c2}}, \boldsymbol{b_2^T}$ are parameters for self-loops.

The same as Eq.3, but have maximum pooling.

$$
\begin{aligned}
h_v^{k+1} = f(\sum_{u \in \mathcal{N}} (\boldsymbol{W}_{l(u,v)}^k h_u^k + \boldsymbol{b}_{l(u,v)}^k)) \\
= \max\{(\boldsymbol{W}_{l(u,v)}^k h_u^k + \boldsymbol{b}_{l(u,v)}^k)|u \in \mathcal{N}\} \\
= \max\{\boldsymbol{W_{c1}V_{w_1}} + \boldsymbol{b_1^T}, \boldsymbol{W_{c3}V_{w_3}} + \boldsymbol{b_3^T}, \boldsymbol{W_{c2}V_w} + \boldsymbol{b_2^T}| \\
(w_1, w), (w, w_3) \in E\} \\
= \max\{\max\{\boldsymbol{W_{c1}V_{w_1}} + \boldsymbol{b_1^T}|\forall w_1\}, \\
\boldsymbol{W_{c2}V_w} + \boldsymbol{b_2^T}, \max\{\boldsymbol{W_{c3}V_{w_3}} + \boldsymbol{b_3^T}|\forall w_3\}\}
\end{aligned} \tag{4}
$$

References

Vashishth, S.; Dasgupta, S. S.; Ray, S. N.; and Talukdar, P. 2018. Dating documents using graph convolution networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, 1605–1615.