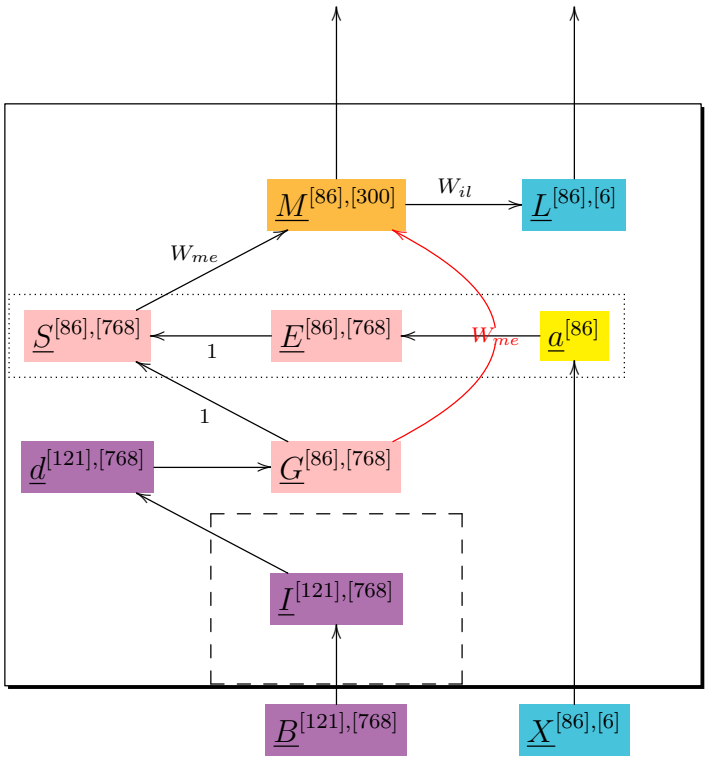


Figure 1: Sax bnet. 2 copies of dashed box are connected in series. 5 copies (5 depths) of plain box are connected in series. However, in the first of those 5 plain box copies, the dotted box is omitted and node $\underline{G}$ feeds directly into node $\underline{M}$ (indicated by red arrow). We display the tensor shape superscripts in the PyTorch L2R order. All tensor shape superscripts have been simplified by omitting a $[s_{ba}]$ from their left side, where $s_{ba} = 24$ is the batch size.

$\underline{a}^{[86]}$ : `ll_greedy_ilabel`
$\underline{B}^{[121],[768]}$ : `lll_hidstate`
$\underline{d}^{[121],[768]}$ : `lll_hidstate`
$\underline{E}^{[86],[768]}$ : `lll_pred_code`
$\underline{G}^{[86],[768]}$ : `lll_word_hidstate`
$\underline{I}^{[121],[768]}$ : `lll_hidstate`
$\underline{L}^{[86],[6]}$ : `lll_word_score`
$\underline{M}^{[86],[300]}$ : `lll_word_hidstate`
$\underline{S}^{[86],[768]}$ : `lll_word_hidstate`
$\underline{X}^{[86],[6]}$ : `lll_word_score`

$$a^{[86]} = \text{argmax}(X^{[86],[6]}; dim = -1) \quad : \texttt{ll\_greedy\_ilabel} \tag{1a}$$

$$B^{[121],[768]} = \mathrm{BERT}() \quad : \texttt{lll\_hidstate} \tag{1b}$$

$$d^{[121],[768]} = \mathrm{dropout}(I^{[121],[768]}) \quad : \texttt{lll\_hidstate} \tag{1c}$$

$$E^{[86],[768]} = \mathrm{embedding}(a^{[86]}) \quad : \texttt{lll\_pred\_code} \tag{1d}$$

$$G^{[86],[768]} = \mathrm{gather}(d^{[121],[768]}; dim = -2) \quad : \texttt{lll\_word\_hidstate} \tag{1e}$$

$$I^{[121],[768]} = \left[B^{[121],[768]} \mathbb{1}(depth = 0) M^{[86],[300]} \mathbb{1}(depth > 0)\right] \quad : \texttt{lll\_hidstate} \tag{1f}$$

$$L^{[86],[6]} = M^{[86],[300]} W_{il}^{[300],[6]} \quad : \texttt{lll\_word\_score} \tag{1g}$$

$$M^{[86],[300]} = \left[G^{[86],[768]} \mathbb{1}(depth = 0) + S^{[86],[768]} \mathbb{1}(depth > 0)\right] W_{me}^{[768],[300]} \quad : \texttt{lll\_word\_hidstate} \tag{1h}$$

$$S^{[86],[768]} = E^{[86],[768]} + G^{[86],[768]} \quad : \texttt{lll\_word\_hidstate} \tag{1i}$$

$$X^{[86],[6]} = L^{[86],[6]} \mathbb{1}(depth > 0) \quad : \texttt{lll\_word\_score} \tag{1j}$$