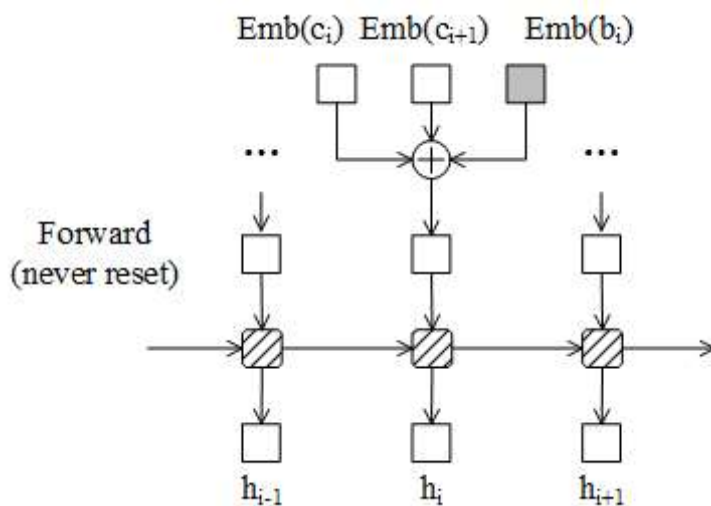
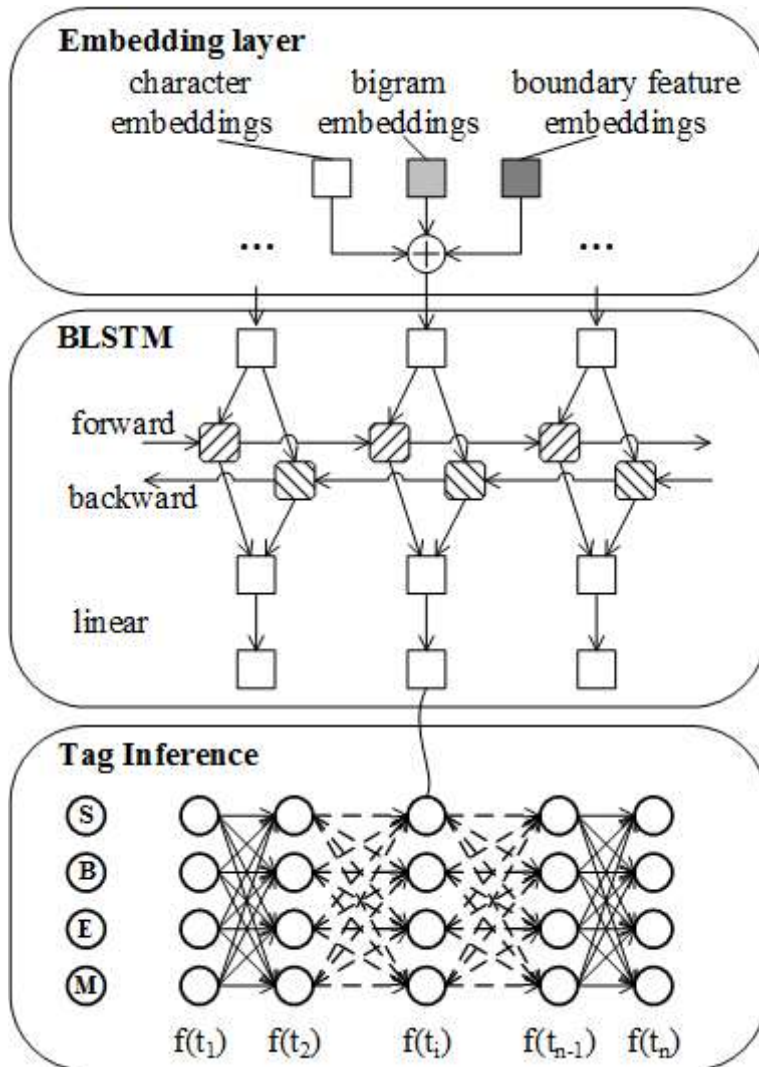
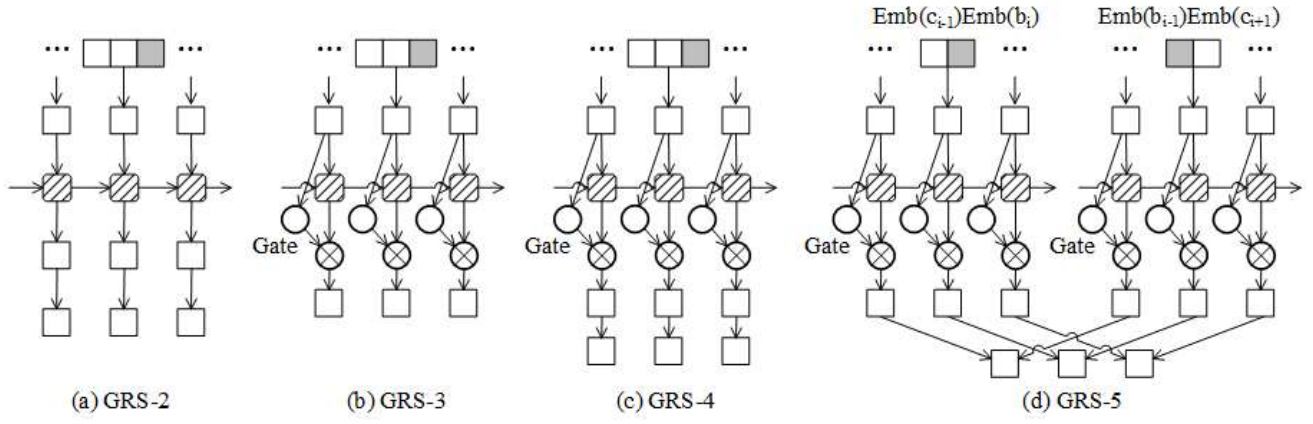


NLP with RNN

Bidirectional LSTM

Use Bidirectional LSTM architecture to get more information about context of the center word





ELMo(Embeddings from Language Models)

For a L – *layer* BiLSTM language model, the ELMo is trained like these

$$\begin{aligned}
 R_k &= \{x_k^{LM}, h_{k,j}^{\leftarrow LM}, h_{k,j}^{\rightarrow LM} | j = 1, \dots, L\} \\
 &= \{h_{k,j}^{LM} | j = 0, \dots, L\} \\
 ELMo_k^{task} &= E(R_k; \Theta^{task}) \\
 &= \gamma^{task} \sum_{j=0}^L s_j^{task} h_k^{LM}
 \end{aligned}$$

s_j^{task} is softmax-normalized weights, γ^{task} is the weight to scale the entire ELMo vector.

while the *Objective* function is

$$\sum_{k=1}^N (\log p(t_k | t_1, \dots, t_{k-1}) + \log p(t_k | t_N, \dots, t_{k+1}))$$

Reference

- [1] Addressing Domain Adaptation for Chinese Word Segmentation with Global Recurrent Structure
- [2] Deep contextualized word representations
- [3] <http://www.wildml.com/category/neural-networks/recurrent-neural-networks/>