

Corpus

$k = 1000$

Sentence containing jiā

$Sent_k^{w_i}$

BERT-base-Chinese (n, 768, 12)

Last layer

Sentence representation (n, 768, 1)

**Token representation (1, 768, 1)**

$e_k^{w_i}$

**Sense**

$$s_j = \arg \max_{s_j} \frac{e^{w_i s_j} \cdot e_k^{w_i}}{\|e^{w_i s_j}\|_2 \|e_k^{w_i}\|_2}$$

CWN

$j = 18$

3 example sentences containing jiā

$\{Sent_1^{w_i s_j}, Sent_2^{w_i s_j}, \dots, Sent_n^{w_i s_j}\}$

3 sentence representations (n, 768, 1)

3 token representations (1, 768, 1)

Averaging

$\{e_1^{w_i s_j}, e_2^{w_i s_j}, \dots, e_n^{w_i s_j}\}$

**Sense representation (1, 768, 1)**

$e^{w_i s_j}$

(Hu et al., 2019)