不代表每个预训练模型的最佳性能

NER（bert+crf）seq\_len=32 epoch=5

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 预训练模型 | batch | micro-f1 | macro-f1 | f1(##WordPiece) | f1(B-NP/I-NP) |
| **chinese\_wwm\_ext\_L-12\_H-768\_A-12** | 40 | 93.0 | 88.0 | 63.0 | 77.0 |
| **chinese\_rbt3\_L-3\_H-768\_A-12** | 40 | 93.1 | 88.0 | 61.0 | 79.0 |
| **chinese\_rbt4\_L-4\_H-768\_A-12** | 40 | 92.0 | 87.0 | 62.0 | 75.0 |
| **chinese\_rbt6\_L-6\_H-768\_A-12** | 40 | 93.0 | 88.0 | 61.0 | 77.0 |
| **chinese\_rbtl3\_L-3\_H-1024\_A-16** | 40 | 92.0 | 88.0 | 64.0 | 74.0 |

SIM（bert）seq\_len=64 epoch=5

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 预训练模型 | batch | accuracy | precision | recall | macro-f1 |
| **chinese\_wwm\_ext\_L-12\_H-768\_A-12** | 40 | 86.0 | 44.3 | 2.9 | 49.0 |
| **chinese\_rbt3\_L-3\_H-768\_A-12** | 40 | 93.5 | 78.3 | 73.1 | 85.9 |
| **chinese\_rbt4\_L-4\_H-768\_A-12** | 40 | 93.8 | 79.2 | 74.9 | 86.7 |
| **chinese\_rbt6\_L-6\_H-768\_A-12** | 40 | 96.5 | 86.4 | 89.1 | 92.9 |
| **chinese\_rbtl3\_L-3\_H-1024\_A-16** | 40 | 95.5 | 82.1 | 86.6 | 90.9 |