**Subword experiment result**

**[1epoch experiment to select the proper hyperparameter (grad\_div)]**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **model** | **seg1** | | **seg2** | | **seg3** | | **seg4** | | **seg5** | |
| **syn** | **sem** | **syn** | **sem** | **syn** | **sem** | **syn** | **sem** | **syn** | **sem** |
| seq35+div32 | 45.0 | 9.3 | 50.1 | 19.6 | 47.5 | 25.5 | 53.2 | 36.2 | 53.9 | 43.5 |
| seq35+div64 | 55.3 | 17.4 | 58.5 | 22.5 | 57.1 | 30.4 | 60.5 | 41.1 | 61.2 | 44.9 |
| seq35+div128 | 61.8 | 15.6 | 63.5 | 26.1 | 62.1 | 28.3 | 65.6 | 35.0 | 66.6 | 37.2 |
| ~~reverse~~ | ~~56.9~~ | ~~13.6~~ | ~~-~~ | ~~-~~ | ~~57.9~~ | ~~27.9~~ | ~~61.1~~ | ~~37.2~~ | ~~61.6~~ | ~~40.5~~ |
| whole+64div | 54.2 | 21.9 | 54.9 | 30.4 | 56.4 | 39.7 | 58.1 | 43.3 | 58.9 | 46.6 |
| whole+128div | - | - | 63.5 | 30.6 | 64.4 | 32.8 | 64.9 | 36.4 | 65.6 | 37.5 |

* ‘reverse’ model이 seq35 model 및 whole 모델 보다도 성능이 낮다.
* seq model보다 whole 모델이 semantic 부분에서 좀 더 학습을 잘 하는 것 같다. Sequence를 나누면서 생긴 양 끝 단에서 context를 만들 때 생긴 noise가 제거되어 semantic한 부분에 대해서 더 잘 학습하는 것 같다.

**[Whole results]**

|  |  |  |
| --- | --- | --- |
| seg25 | div64 | div128 |
| whole | syntactic accuracy: 59.1 [%]  semantic accuracy: 49.3 [%] | syntactic accuracy: 62.5 [%]  semantic accuracy: 48.7 [%] |