# Progress of Neural Machine Translation with Memory Network

Yang FENG

Jan 3, 2016

# 1 Baseline

Improve the baseline by

1. rnn to bidrectional_rnn;

2. SGD optimizer to Adadelta optimization;

3. initializing the initial attention state;

4. seperating hidden_edim and hidden_units (in the baseline, they share a common parameter);

5. changing the frequency of saving models.

To do:
change one-best greedy search to **beam search**.

Cannot get reasonable result of baseline, shown in Table 1. Investigated in terms of

1. output order chaos;

2. file format.

Fix this problem by using greater learning rate of 0.5, and results are shown in Table 2.

# 2 NMT+MN

## 2.1 Adding attended lexical translation to the decoder

| System | Speed (per epoch) | Dev (BLEU4 ) | Test (BLEU4) |
|---|---|---|---|
| nmt | 84s | 10.3 | 11.1 (49w epch) |
| nmt$^+$ | 108s | 13.0 | 15.4 (26w epch) |
| nmt$^+$+mn | 180s | 11.1 | 12.5 (30w epch) |

| System | Speed (per epoch) | Dev (BLEU4 ) | Test (BLEU4) |
|---|---|---|---|
| nmt | 84s | 28.3 | 30.0 (24600 step) |