Context-Aware Smoothing for Neural Machine Translation

Kehai Chen¹, Rui Wang², Masao Utiyama², Eiichiro Sumita² and Tiejun Zhao¹

¹Harbin Institute of Technology, China
²National Institute of Information and Communications Technology, Japan
Content

• Motivation
• Related Works
• Context-Aware Representation
• NMT with Context-Aware Smoothing
• Experimental Results
• Conclusion
## Motivation-1: polysemic words

| Src1  | 他们 想 通过 打 比赛 来 解决 矛盾 |
|-------|--------------------------------|
| Trg1  | They want to solve the dispute by playing the game |
| Src2  | 他们 正在 因为 争执 而 打 对方 |
| Trg2  | They are beating each other for a dispute |

### Two bilingual parallel sentence pairs

The lexicon semantic depends on its specific context.
Motivation-1: Enhancing word representation for polysemy words

Learn specific-sentence word representation $\mathbf{v}_j$

$$h_j = f_{\text{enc}}(\mathbf{v}_j, h_{j-1})$$

Better source representation

... ...

Better target translation

Two bilingual parallel sentence pairs

| Src1  | 他们 想 通过 打 比赛 来 解决 矛盾 |
|-------|---------------------------------|
| (pinyin) | tamen xiang tongguo da bisai lai jiejue maodun |

| Trg1  | They want to solve the dispute by playing the game |
|-------|---------------------------------------------------|
|       | $\mathbf{v} \xrightarrow{\text{playing}}$        |
| da    | $\mathbf{v}_d$                                    |

| Src2  | 他 们 正在 因为 争执 而 打 对方 |
|-------|--------------------------------|
| (pinyin) | tamen zhengzai yinwei zhengzhi er da duifang |

| Trg2  | They are beating each other for a dispute |
|-------|------------------------------------------|
|       | $\mathbf{v} \xrightarrow{\text{beating}}$ |
| da    | $\mathbf{v}_d$                                     |

The lexicon semantic depends on its specific context
Motivation-2: OOV

x₁, x₂, x₃, x₄, x₅ → h₁, h₂, h₃, h₄, h₅ → v₁, v₂, v₃, v₄, v₅

Word embedding layer

Encoder layer

Decoder layer

y₁, y₂, y₃, y₄, y₅ → c₁, c₂, c₃, c₄, c₅

Attention α

Encoder-Decoder NMT

11/28/2017
Motivation-2: OOV

- The source sentence includes a OOV

Single vector $v_u$ represents all OOVs
Motivation-2: OOV

- The source sentence includes an OOV
  - Single vector $v_u$ represents all OOVs

- Breaking the structure of the sentence;
- Pooling source representation;
- ... ...
- Affecting translation prediction of target word.

These gray parts indicate the parameters of NMT which are affected by the OOV.
Related Works

• Translation Granularity for NMT
  ---Smaller Translation Granularity: Word, Sub-word (BPE), Character for OOV.
    Sennrich et al. (2016), Costa-jussa and Fonollosa (2016), and Li et al. (2016), ... ...

• Source representation for NMT
  ---RNN or CNN-based Encoder: learning source representation over the sequence of fixed word vectors.
    Bahdanau et al. (2015), Sutskever et al. (2014), ... ...
Related Works

• Translation Granularity for NMT
  ---Smaller Translation Granularity: Word, Sub-word (BPE), Character for OOV.
  Sennrich et al. (2016), Costa-jussa and Fonollosa (2016), and Li et al. (2016), ... ...

• Source representation for NMT
  ---RNN or CNN-based Encoder: learning source representation over the sequence of fixed word vectors.
  Bahdanau et al. (2015), Sutskever et al. (2014), ... ...

• This work focus on enhancing word embedding layer.
  ---Learning a specific-sentence representation for polysemy or OOV word by its context words.
  ---Offering context-aware representation enhances word embedding layer, thereby improving translations (though RNN Encoder can capture word context).
Context-Aware Representation

If there is an OOV “unk” (or polysemy word) in the sentence:

\[ x_1 \ x_2 \ x_3 \ x_4 \ unk \ x_6 \ x_7 \ x_8 \ x_9 \]

When one understands natural language sentence intuitively, especially including OOV or polysemy word, one often inferences the meaning of these words depending on its context words.

\[ v_1 \ v_2 \ v_3 \ v_4 \ \text{unk} \ v_6 \ v_7 \ v_8 \ v_9 \]
Context-Aware Representation

• We define a context $L_j$ for source word $x_j$ in a fixed size window $2n$:

$$L_j = x_{j-n}, \ldots, x_{j-1}, \ x_{j+1}, \ldots, \ x_{j+n}$$

Historical $n$ words  
Future $n$ words
We define a context $L_j$ for source word $x_j$ in a fixed size window $2n$:

$$L_j = x_{j-n}, \ldots, x_{j-1}, \ x_{j+1}, \ldots, x_{j+n}$$

Historical $n$ words  Future $n$ words

Take $x_5$ as an example, its context $L_5$ follows ($n=2$):

$x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ x_6 \ x_7 \ \ldots \ \ x_J$

$L_5 = x_3, x_4, x_6, x_7$
Feedforward Context-of-Words Model (FCWM)

Output layer:

$$V_{L_j} = \sigma(W_1L_j + b_1)$$

Concatenation:

$$L_j = [v_{j-n}: \ldots : v_{j-1} : v_{j+1} : \ldots : v_{j+n}]$$

Input layer:

$$L_j = v_{j-n}, \ldots, v_{j-1}, v_{j+1}, \ldots, v_{j+n}$$

Context words $L_j$ of $x_j$:

$$L_j = x_{j-n}, \ldots, x_{j-1}, x_{j+1}, \ldots, x_{j+n}$$

$$V_{L_j} = \varphi_1 (L_j; \theta_1)$$
Context-Aware Representation

Feedforward Context-of-Words Model (FCWM)

Output layer:

\[ V_{Lj} = \sigma(W_1 L_j + b_1) \]

Concatenation:

\[ L_j = [v_{j-n}; \ldots; v_{j-1}; v_{j+1}; \ldots; v_{j+n}] \]

Input layer:

\[ L_j = v_{j-n}, \ldots, v_{j-1}, v_{j+1}, \ldots, v_{j+n} \]

Context words \( L_j \) of \( x_j \):

\[ L_j = x_{j-n}, \ldots, x_{j-1}, x_{j+1}, \ldots, x_{j+n} \]

\[ V_{Lj} = \varphi_1 (L_j; \theta_1) \]

Convolutional Context-of-Words Model (CCWM)

Non-linear output layer:

\[ V_{Lj} = \sigma(W_3 (\text{ave}(\sum_{l=1}^{2n-k+1} P_l)) + b_3) \]

Pooling layer:

\[ P = \max[P_1, \ldots, P_{2n-k+1}] \]

\[ P_l = \max[L_{2l-1}, L_{2l}] \]

Convolution layer:

\[ L = [L_1, \ldots, L_{2n-k+1}] \]

Input layer:

\[ L_j = \psi(W_2 M + b_2) \]

Context words \( L_j \) of \( x_j \):

\[ L_j = x_{j-n}, \ldots, x_{j-1}, x_{j+1}, \ldots, x_{j+n} \]

\[ V_{Lj} = \varphi_2 (L_j; \theta_2) \]
NMT for OOV Smoothing

Standard NMT:
\[ p(y_i|y_{<i}, x) = g(v_{y_{i-1}}, s_i, c_i) \]

This work:
\[ p(y_i|y_{<i}, x) = \begin{cases} g(v_{y_{i-1}}, s_i, c_i), & y_{i-1} \in V_t \\ g(\varphi_d(L_{y_{i-1}}), s_i, c_i), & y_{i-1} \notin V_t \end{cases} \]

• CARNMT-Enc

This work:
\[ h_j = \begin{cases} f_{\text{enc}}(v_j, h_{j-1}), & x_j \in V_s \\ f_{\text{enc}}(\varphi_e(L_{x_j}), h_{j-1}), & x_j \notin V_s \end{cases} \]
NMT for OOV Smoothing

Standard NMT:
\[ p(y_i | y_{<i}, x) = g(v_{y_{i-1}}, s_i, c_i) \]

This work:
\[ p(y_i | y_{<i}, x) = \begin{cases} 
  g(v_{y_{i-1}}, s_i, c_i), & y_{i-1} \in V_t \\
  g(\varphi_d(L_{y_{i-1}}), s_i, c_i), & y_{i-1} \notin V_t 
\end{cases} \]

**CARNMT-Dec**

This work:
\[ h_j = f_{enc}(v_j, h_{j-1}) \]

This work:
\[ h_j = \begin{cases} 
  f_{enc}(v_j, h_{j-1}), & x_j \in V_s \\
  f_{enc}(\varphi_e(L_{x_j}), h_{j-1}), & x_j \notin V_s 
\end{cases} \]
NMT for OOV Smoothing

Standard NMT:
\[ h_j = f_{\text{enc}}(v_j, h_{j-1}) \]

This work:
\[ h_j = \begin{cases} f_{\text{enc}}(v_j, h_{j-1}), & x_j \in V_s \\ f_{\text{enc}}(\varphi_e(L_{x_j}), h_{j-1}), & x_j \notin V_s \end{cases} \]

Standard NMT:
\[ p(y_i | y_{i<i}, x) = g(v_{y_{i-1}}, s_i, c_i) \]

This work:
\[ p(y_i | y_{<i}, x) = \begin{cases} g(v_{y_{i-1}}, s_i, c_i), & y_{i-1} \in V_t \\ g(\varphi_d(L_{y_{i-1}}), s_i, c_i), & y_{i-1} \notin V_t \end{cases} \]
Standard NMT:
\[ p(y_i|y_{i<i}, x) = g(v_{y_{i-1}}, s_i, c_i) \]

This work:
\[ p(y_i|y_{i<i}, x) = g(\phi_d(L_{y_{i-1}}), s_i, c_i) \]
Experimental Settings

- Training data includes 1.42M Chinese-to-English parallel sentence pairs from LDC corpus.
- The NIST 2002 (MT02) and NIST 2003-2008 (MT03-08) datasets are as validation set and test sets, respectively. The Case-insensitive 4-gram NIST BLEU score (Papineni et al., 2002) is as evaluation metric.
- Vocab is 30k; Sentence length is 80; Mini-batch size 80; Word embedding dim is 620; Hidden layer dim is 1000; Dropout on the all layers; Optimizer is Adadelta.
- The baseline includes: Standard Attentional NMT (Bahdanau et al., 2014); Subword-based NMT (Sennrich et al., 2016); Character-based NMT (Costa-jussa and Fonollosa, 2016); Replacing unk with similarity semantic in vocabulary words (Li et al., 2016).
Experimental Results

• Results for Chinese-to-English Translation Task

| System                        | Dev (MT02) | MT03  | MT04  | MT05  | MT06  | MT08  | AVG  |
|-------------------------------|------------|-------|-------|-------|-------|-------|------|
| Moses                         | 33.15      | 31.02 | 33.78 | 30.33 | 29.62 | 23.53 | 29.66|
| Bahdanau et al. (2015)        | 36.42      | 34.22 | 37.11 | 33.02 | 32.69 | 25.38 | 32.48|
| Sennrich et al. (2016)        | 36.89      | 35.39 | 38.24 | 33.73 | 32.74 | 26.22 | 33.26|
| Costa-jussà and Fonollosa (2016) | 35.98  | 34.93 | 37.56 | 33.24 | 32.32 | 26.02 | 32.81|
| Li et al. (2016)              | 36.96      | 35.78 | 38.42 | 34.02 | 33.14 | 26.36 | 33.54|
| CARNMT-Encoder (FCWM)         | 36.78      | 35.56**| 38.14*| 33.69 | 33.13 | 26.16*| 33.34|
| CARNMT-Decoder (FCWM)         | 36.67      | 34.65 | 37.60 | 33.26 | 33.01 | 26.15*| 32.93|
| CARNMT-Both (FCWM)            | 37.36      | 35.43**| 38.34**| 33.43 | 33.47 | 26.86**| 33.50|
| ALLSmooth (FCWM)              | 37.71      | 35.73**| 38.53**| 33.91*| 33.53*| 27.18**| 33.78|
| CARNMT-Encoder (CCWM)         | 37.12      | 35.64**| 38.14*| 33.49 | 33.26*| 26.57**| 33.42|
| CARNMT-Decoder (CCWM)         | 36.33      | 34.56 | 37.43 | 33.24 | 32.96 | 25.86 | 32.81|
| CARNMT-Both (CCWM)            | 37.56      | 35.83**| 38.52**| 33.73 | 33.37**| 27.06**| 33.70|
| ALLSmooth (CCWM)              | 37.69      | 36.23**| 38.89**| 34.69**| 33.83**| 27.94†| 34.32|

• Moses VS NMT --------> Strong baselines
  • CARNMT-Enc/Dec VS Bahdanau et al. (2015) --------> Our method can effectively smooth the negative effect (Motivation 1)
  • CARNMT-Both VS CARNMT-Enc/Dec --------> Source-side smoothing is orthogonal with target-side smoothing (Motivation 1)
  • ALLSmooth VS CARNMT-Both --------> In-vocabulary smoothing is beneficial for NMT (Motivation 2)

11/28/2017
Experimental Results

• Results for Chinese-to-English Translation Task

| System                              | Dev (MT02) | MT03 | MT04 | MT05 | MT06 | MT08 | AVG |
|-------------------------------------|------------|------|------|------|------|------|-----|
| Moses                               | 33.15      | 31.02| 33.78| 30.33| 29.62| 23.53| 29.66|
| Bahdanau et al. (2015)              | 36.42      | 34.22| 37.11| 33.02| 32.69| 25.38| 32.48|
| Sennrich et al. (2016)              | 36.89      | 35.39| 38.24| 33.73| 32.74| 26.22| 33.26|
| Costa-jussà and Fonollosa (2016)    | 35.98      | 34.93| 37.56| 33.24| 32.32| 26.02| 32.81|
| Li et al. (2016)                    | 36.96      | 35.78| 38.42| 34.02| 33.14| 26.36| 33.54|
| CARNMT-Encoder (FCWM)               | 36.78      | 35.56**| 38.14*| 33.69| 33.13| 26.16*| 33.34|
| CARNMT-Decoder (FCWM)               | 36.67      | 34.65| 37.60| 33.26| 33.01| 26.15*| 32.93|
| CARNMT-Both (FCWM)                  | 37.36      | 35.43**| 38.34**| 33.43| 33.47| 26.86**| 33.50|
| ALLSmooth (FCWM)                    | 37.71      | 35.73**| 38.53**| 33.91*| 33.53*| 27.18**| 33.78|
| CARNMT-Encoder (CCWM)               | 37.12      | 35.64**| 38.14*| 33.49| 33.26*| 26.57**| 33.42|
| CARNMT-Decoder (CCWM)               | 36.33      | 34.56| 37.43| 33.24| 32.96| 25.86| 32.81|
| CARNMT-Both (CCWM)                  | 37.56      | 35.83**| 38.52**| 33.73| 33.37**| 27.06**| 33.70|
| ALLSmooth (CCWM)                    | 37.69      | 36.23**| 38.89**| 34.69**| 33.83**| 27.94†| 34.32|

• CARNMT-Enc/Dec VS Bahdanau et al. (2015)
  Our smooth method can relieve the negative effect of OOV effectively, as in Motivation 2
  • CARNMT-Both VS CARNMT-Enc/Dec -----> Source-side smoothing is orthogonal with target-side smoothing (Motivation 1)
  • ALLSmooth VS CARNMT-Both -----> In-vocabulary smoothing is beneficial for NMT (Motivation 2)
Experimental Results

• Results for Chinese-to-English Translation Task

| System                          | Dev (MT02) | MT03  | MT04  | MT05  | MT06  | MT08  | AVG   |
|---------------------------------|------------|-------|-------|-------|-------|-------|-------|
| Moses                           | 33.15      | 31.02 | 33.78 | 30.33 | 29.62 | 23.53 | 29.66 |
| Bahdanau et al. (2015)          | 36.42      | 34.22 | 37.11 | 33.02 | 32.69 | 25.38 | 32.48 |
| Sennrich et al. (2016)          | 36.89      | 35.39 | 38.24 | 33.73 | 32.74 | 26.22 | 33.26 |
| Costa-jussà and Fonollosa (2016)| 35.98      | 34.93 | 37.56 | 33.24 | 32.32 | 26.02 | 32.81 |
| Li et al. (2016)                | 36.96      | 35.78 | 38.42 | 34.02 | 33.14 | 26.36 | 33.54 |
| CARNMT-Encoder (FCWM)           | 36.78      | 35.56**| 38.14*| 33.69 | 33.13 | 26.16*| 33.34 |
| CARNMT-Decoder (FCWM)           | 36.67      | 34.65 | 37.60 | 33.26 | 33.01 | 26.15*| 32.93 |
| CARNMT-Both (FCWM)              | 37.36      | 35.43**| 38.34**| 33.43 | 33.47 | **26.86**| 33.50 |
| ALLSmooth (FCWM)                | 37.71      | 35.73**| **38.53**| 33.91*| **33.53**| **27.18**| 33.78 |
| CARNMT-Encoder (CCWM)           | 37.12      | 35.64**| 38.14*| 33.49 | 33.26*| **26.57**| 33.42 |
| CARNMT-Decoder (CCWM)           | 36.33      | 34.56 | 37.43 | 33.24 | 32.96 | 25.86 | 32.81 |
| CARNMT-Both (CCWM)              | 37.56      | **35.83**| **38.52**| 33.73 | **33.37**| **27.06**| **33.70** |
| ALLSmooth (CCWM)                | 37.69      | **36.23**| **38.89**| **34.69**| **33.83**| **27.94**| **34.32** |

• CARNMT-Both VS CARNMT-Enc/Dec
  Source-side smoothing is orthogonal with target-side smoothing
# Experimental Results

## Results for Chinese-to-English Translation Task

| System                            | Dev (MT02) | MT03 | MT04  | MT05  | MT06  | MT08  | AVG  |
|-----------------------------------|------------|------|-------|-------|-------|-------|------|
| Moses                             | 33.15      | 31.02| 33.78 | 30.33 | 29.62 | 23.53 | 29.66|
| Bahdanau et al. (2015)            | 36.42      | 34.22| 37.11 | 33.02 | 32.69 | 25.38 | 32.48|
| Sennrich et al. (2016)            | 36.89      | 35.39| 38.24 | 33.73 | 32.74 | 26.22 | 33.26|
| Costa-jussà and Fonollosa (2016)  | 35.98      | 34.93| 37.56 | 33.24 | 32.32 | 26.02 | 32.81|
| Li et al. (2016)                  | 36.96      | 35.78| 38.42 | 34.02 | 33.14 | 26.36 | 33.54|
| CARNMT-Encoder (FCWM)             | 36.78      | 35.56| 38.14*| 33.69 | 33.13 | 26.16*| 33.34|
| CARNMT-Decoder (FCWM)             | 36.67      | 34.65| 37.60 | 33.26 | 33.01 | 26.15*| 32.93|
| CARNMT-Both (FCWM)                | 37.36      | 35.43*| 38.34**| 33.43 | 33.47 | 26.86**| 33.50|
| ALLSmooth (FCWM)                  | 37.71      | 35.73**| 38.53**| 33.91*| 33.53*| 27.18**| 33.78|
| CARNMT-Encoder (CCWM)             | 37.12      | 35.64**| 38.14*| 33.49 | 33.26*| 26.57**| 33.42|
| CARNMT-Decoder (CCWM)             | 36.33      | 34.56| 37.43 | 33.24 | 32.96 | 25.86 | 32.81|
| CARNMT-Both (CCWM)                | 37.56      | 35.83**| 38.52**| 33.73 | 33.37**| 27.06**| 33.70|
| ALLSmooth (CCWM)                  | 37.69      | **36.23**| **38.89**| **34.69**| **33.83**| **27.94**| **34.32**|

**ALLSmooth VS CARNMT-Both**

In-vocabulary smoothing is also beneficial for NMT (Motivation 1)
Experimental Results

• Results for Chinese-to-English Translation Task

| System                                      | Dev (MT02) | MT03 | MT04 | MT05 | MT06 | MT08 | AVG  |
|---------------------------------------------|------------|------|------|------|------|------|------|
| Moses                                       | 33.15      | 31.02| 33.78| 30.33| 29.62| 23.53| 29.66|
| Bahdanau et al. (2015)                      | 36.42      | 34.22| 37.11| 33.02| 32.69| 25.38| 32.48|
| Sennrich et al. (2016)                      | 36.89      | 35.39| 38.24| 33.73| 32.74| 26.22| 33.26|
| Costa-jussà and Fonollosa (2016)            | 35.98      | 34.93| 37.56| 33.24| 32.32| 26.02| 32.81|
| Li et al. (2016)                            | 36.96      | 35.78| 38.42| 34.02| 33.14| 26.36| 33.54|
| CARNMT-Encoder (FCWM)                       | 36.78      | 35.56**| 38.14*| 33.69| 33.13| 26.16*| 33.34|
| CARNMT-Decoder (FCWM)                       | 36.67      | 34.65| 37.60| 33.26| 33.01| 26.15*| 32.93|
| CARNMT-Both (FCWM)                          | 37.36      | 35.43**| 38.34**| 33.43| 33.47| 26.86**| 33.50|
| ALLSmooth (FCWM)                            | 37.71      | 35.73**| **38.53**| 33.91*| **33.53**| **27.18**| **33.78**|
| CARNMT-Encoder (CCWM)                       | 37.12      | 35.64**| 38.14*| 33.49| 33.26*| 26.57**| 34.42|
| CARNMT-Decoder (CCWM)                       | 36.33      | 34.56| 37.43| 33.24| 32.96| 25.86| 32.81|
| CARNMT-Both (CCWM)                          | 37.56      | **35.83**| **38.52**| 33.73| **33.37**| **27.06**| **33.70**|
| ALLSmooth (CCWM)                            | 37.69      | **36.23**| **38.89**| **34.69**| **33.83**| **27.94**| **34.32**|

• FCWM VS CCWM

The CCWM learns the context semantic representation directly for smoothing word vector, while the FCWM predicts semantic representation of word depending on its context.
Experimental Results

- Translation Qualities for Sentences with Different Numbers of OOV

- The number of OOV = 0
  - ALLSmooth is better than the baseline Bahdanau et al. (2015).
  - Both of CARNMT-Enc/Dec are similar to baseline Bahdanau et al. (2015).
- With the increasing in the number of OOVs
  - The gap between our methods and other methods (except PBSMT) become larger, especially when more than five.
  - When the number of OOV is more than seven
    - PBSMT is better than all NMT models

The number of sentences: 2306, 1827, 1121, 678, 391, 215, 123, 59, 37, 24, 29
Experimental Results

- Translation Qualities for Sentences with Different Numbers of OOV

- The number of OOV = 0
  - ALLSmooth is better than the baseline Bahdanau et al. (2015).
  - Both of CARNMT-Enc/Dec are similar to baseline Bahdanau et al. (2015).

- With the increasing in the number of OOVs
  - The gap between our methods and other methods (except PBSMT) become larger, especially when more than five.
  - When the number of OOV is more than seven
    - PBSMT is better than all NMT models

The number of sentences: 2306 1827 1121 678 391 215 123 59 37 24 29

11/28/2017
Experimental Results

- Translation Qualities for Sentences with Different Numbers of OOV

- The number of OOV = 0
  - ALLSmooth is better than the baseline Bahdanau et al. (2015).
  - Both of CARNMT-Enc/Dec are similar to baseline Bahdanau et al. (2015).

- With the increasing in the number of OOVs
  - The gap between our methods and other methods (except PBSMT) become larger, especially when more than five.
  - When the number of OOV is more than seven
    - PBSMT is better than all NMT models

The number of sentences

| 2306 | 1827 | 1121 | 678  | 391  | 215  | 123  | 59   | 37   | 24   | 29   |
Experimental Results

- Translation Qualities for Sentences with Different Numbers of OOV

- The number of OOV = 0
  - ALLSmooth is better than the baseline Bahdanau et al. (2015).
  - Both of CARNMT-Enc/Dec are similar to baseline Bahdanau et al. (2015).

- With the increasing in the number of OOVs
  - The gap between our methods and other methods (except PBSMT) become larger, especially when more than five.

- When the number of OOV is more than seven
  - PBSMT is better than all NMT models

The number of sentences

2306 1827 1121 678 391 215 123 59 37 24 29

The number of sentences
Experimental Results

- Translation Qualities for Sentences with Different Numbers of OOV

\[ \text{SRC: 用好 这个 战略 机遇期 (OOV), 力争 有所 作为, 必须 把 发展 科学 技术 放在 更加 重要, 更加 突出 的 位置} \]
\[ \text{(pinyin) yonghao zhege zhanlue jiyuqi , lizheng yousuo zuowei, bixu ba fazhan kexue jishu} \]
\[ \text{fangzai gengjia zhongyao, gengjia tuchu de weizhi} \]

\[ \text{Bahdanau et al. (2015): to make good use of this strategy, we should strive for the development of} \]
\[ \text{science and technology, and must put the development of science and technology into an even more} \]
\[ \text{important and prominent position} \]

\[ \text{This work: in making good use of this strategic plan and striving to accomplish something, it is necessary} \]
\[ \text{to place the development of science and technology in a more important and more prominent position} \]

\[ \text{Ref: to well use this strategic period of opportunity and strive to accomplish some achievements, the development of} \]
\[ \text{science and technology should be placed in a more prior and prominent position} \]

- The negative effect of OOV exists in NMT
  - The OOV “jiyuqi” itself is not translated.
  - The phrase “lizheng yousuo zuowei” (the red part in English) is not translated.
- Smoothing the negative effect of OOV
  - Obtaining the translation “striving to accomplish something” of “lizheng yousuo zuowei”.
Conclusion

• Experimental results showed that the negative effect of OOV decreased the translation performance of NMT, and the existing RNN encoder can not adequately address the problem.

• The learned CAR was integrated into the Encoder to smooth word representation, and thus enhanced the Decoder of NMT.

• Experimental results showed that the proposed method can greatly alleviate the negative effect of OOV and enhance word representation of in-vocabulary words, thus improving the translations.
Q&A
Thanks