1 Neural Machine Translation with RNNs

(g) The purpose is to pay no attention to the padding tokens by setting their values to negative infinity.

(j) Comparisons of different attention mechanisms.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Dot Product | Multiplicative | Additive |
| Advantage | Fast to compute | Can fit different dimensions of the hidden layers of the encoder and decoder | High complexity gives it ability to learn more relationships |
| Disadvantage | Only suitable for same dimensions of the hidden layers of the encoder and decoder | Less parameters, can learn limited knowledge | High complexity needs more data to learn the parameters and takes longer time to converge |

2 Analyzing NMT Systems

(c)

i.

c1 analysis:

|  |  |  |  |
| --- | --- | --- | --- |
| 1-gram | Count | Max-count | Count clip |
| the | 1 | 0 | 0 |
| love | 1 | 1 | 1 |
| can | 1 | 1 | 1 |
| always | 1 | 1 | 1 |
| do | 1 | 0 | 0 |

p1 = 3/5 = 0.6

|  |  |  |  |
| --- | --- | --- | --- |
| 2-gram | Count | Max-count | Count clip |
| the love | 1 | 0 | 0 |
| love can | 1 | 1 | 1 |
| can always | 1 | 1 | 1 |
| always do | 1 | 0 | 0 |

P2 = 2/4 = 0.5

BLEU(c1) = 0.5477

c2 analysis

|  |  |  |  |
| --- | --- | --- | --- |
| 1-gram | Count | Max-count | Count clip |
| love | 1 | 1 | 1 |
| can | 1 | 1 | 1 |
| make | 1 | 0 | 0 |
| anything | 1 | 1 | 1 |
| possible | 1 | 1 | 1 |

p1 = 4/5 = 0.8

|  |  |  |  |
| --- | --- | --- | --- |
| 2-gram | Count | Max-count | Count clip |
| love can | 1 | 1 | 1 |
| can make | 1 | 0 | 0 |
| make anything | 1 | 0 | 0 |
| anything possible | 1 | 1 | 1 |

p2 = 2/4 = 0.5

BLUE(c2) = 0.6325

ii.

c1 analysis:

|  |  |  |  |
| --- | --- | --- | --- |
| 1-gram | Count | Max-count | Count clip |
| the | 1 | 0 | 0 |
| love | 1 | 1 | 1 |
| can | 1 | 1 | 1 |
| always | 1 | 1 | 1 |
| do | 1 | 0 | 0 |

p1 = 3/5 = 0.6

|  |  |  |  |
| --- | --- | --- | --- |
| 2-gram | Count | Max-count | Count clip |
| the love | 1 | 0 | 0 |
| love can | 1 | 1 | 1 |
| can always | 1 | 1 | 1 |
| always do | 1 | 0 | 0 |

P2 = 2/4 = 0.5

BLEU(c1) = 0.5477

c2 analysis

|  |  |  |  |
| --- | --- | --- | --- |
| 1-gram | Count | Max-count | Count clip |
| love | 1 | 1 | 1 |
| can | 1 | 1 | 1 |
| make | 1 | 0 | 0 |
| anything | 1 | 0 | 0 |
| possible | 1 | 0 | 0 |

p1 = 2/5 = 0.4

|  |  |  |  |
| --- | --- | --- | --- |
| 2-gram | Count | Max-count | Count clip |
| love can | 1 | 1 | 1 |
| can make | 1 | 0 | 0 |
| make anything | 1 | 0 | 0 |
| anything possible | 1 | 0 | 0 |

p2 = 1/4 = 0.25

BLUE(c2) = 0.1900

The result is bad.

iii. A same Meaning can be expressed in different ways. Different people have different speaking habits.

iv.

Advantages: Fast to compute; Can evaluate the quality of a translation in some degree.

Disadvantages: Can’t judge if a translation is fluent; Can’t distinguish words with the same meaning.