

Sequence to sequence models

Bleu score (optional)

Evaluating machine translation

French: Le chat est sur le tapis.

Precision: $\frac{7}{2}$

Dilipped exolution understudy

Reference 1: The cat is on the mat.

Reference 2: There is a cat on the mat.

MT output: the the the the the the.

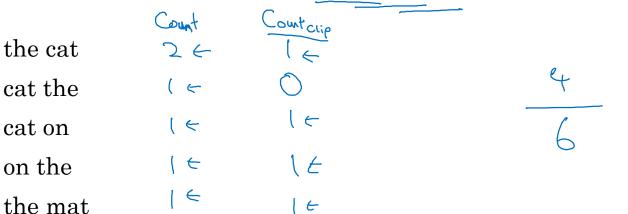
[Papineni et. al., 2002. Bleu: A method for automatic evaluation of machine translation]

Bleu score on bigrams (M) (M) (M)

Example: Reference 1: The cat is on the mat.

Reference 2: There is a cat on the mat.

MT output: The cat the cat on the mat. \leftarrow



Bleu score on unigrams

Example: Reference 1: The cat is on the mat.

Reference 2: There is a cat on the mat.

Therefore 2. There is a car on the man

→ MT output: The cat the cat on the mat. _count city (unigram) unigram€\$ count (unigram) Wigcom

[Papineni et. al., 2002. Bleu: A method for automatic evaluation of machine translation]

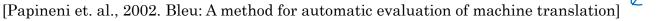
Andrew Ng

Bleu details

$$p_n$$
 = Bleu score on n-grams only

Combined Bleu score:
$$\mathbb{R}^{p_{k}} \exp\left(\frac{1}{T} \sum_{n=1}^{q_{k}} \mathbb{P}_{n}\right)$$

$$BP = \begin{cases} 1 & \text{if MT_output_length} > \text{reference_output_length} \\ \exp(1 - \text{MT_output_length}) & \text{otherwise} \end{cases}$$



P1, P2, P3, P4