Bleu score

Une image contenant texte, Police, capture d’écran, ligne

Description générée automatiquement

Reference 🡪 true label

System output 🡪 pred label

Bleu compares how similar they are

(step 1)

Split reference to tokens:

You – are – four – minutes – too – slow

(step 2) compute n grams

* 1 gram: [“you”, “are”, “four”, “minutes”, “too”, “slow”]
* 2 gram: [“you are”, “are four”, “four minutes”, “minutes too”, “too slow”]
* 3 gram: [“you are four”, “are four minutes”, “four minutes too”, “minutes too slow”]
* 4 gram: [“you are four minutes”, “are four minutes too”, “four minutes too slow”]

For each n gram compute weather the they are in the reference and candidate sentences

Une image contenant texte, capture d’écran, Police, nombre

Description générée automatiquement

The score of ngram 1 is:

4 🡪 sum of the last col

5 🡪 number of 1 grams in the candidate output (here 5)

(for 2 gram it will be 4 … etc )





Number of words in reference = 6

Number of words in candidate = 5

Final score:

Une image contenant texte, Police, Bleu électrique, capture d’écran

Description générée automatiquement