N-ROUGE Metric

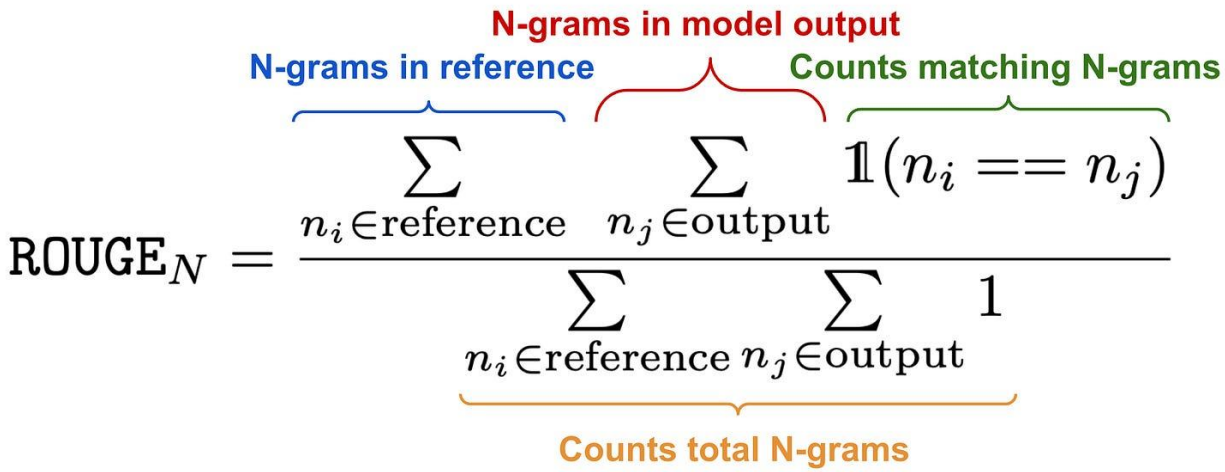
ROUGE-N specifically measures n-gram overlap between the generated and reference text.

ROUGE-N is the **recall** of n-grams between a candidate (generated) and reference (ground-truth) text.

Variants:

- **ROUGE-1** : Unigram overlap (focuses on content)
- **ROUGE-2** : Bigram overlap (considers fluency/coherence)
- **ROUGE-L** : Longest Common Subsequence
- **ROUGE-S** : Skip-bigram (gapped bigrams)

Formula:



Example (ROUGE-1):

Let's say:

- **Reference:** "The cat sat on the mat"
- **Candidate:** "The cat lay on the mat"

Unigrams (n=1):

- Reference unigrams: ["The", "cat", "sat", "on", "the", "mat"]
- Candidate unigrams: ["The", "cat", "lay", "on", "the", "mat"]

Overlap: ["The", "cat", "on", "the", "mat"] → 5 out of 6

ROUGE-1 = 5/6 ≈ 0.83

Recall vs. Precision

- **ROUGE-N** emphasizes **recall**:

  How much of the reference text is captured in the generated output

- **BLEU** emphasizes **precision**:

  How much of the generated output matches the reference