

BLEU scores are a way to check how good a machine translation is by comparing it to a translation done by a person. It looks at small groups of words (n-grams) in both the machine's translation and the human translation to see how similar they are.

Here's how it works simply:

1. **Count Matching Words:** It counts how many small groups of words in the machine's translation also appear in the human translation.
2. **Penalty for Short Translations:** If the machine's translation is shorter than the human one, it gets a lower score. This is to make sure that shorter translations don't just get a good score because they are short.

BLEU scores are useful because they are easy to use and understand, but they aren't perfect. They don't always catch if the translation sounds natural or follows good grammar.

Sources:

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). BLEU: a Method for Automatic Evaluation of Machine Translation. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), pages 311-318.

Related posts:

1. Nvidia toolkit
2. Impact Factor
3. Key Challenges in Natural Language Processing Explained
4. What is NLP ?
5. SciSpace Alternatives: Top Tools for Researchers