Machine Learning Analysis

There are four main methods used to assess the accuracy of closed captions made with machine learning. They are called natural language methods because they each measure how close the captions are to real human writing. The most common metrics are:

- BLEU measures the accuracy of machine-generated text by analyzing whole sentences.
- METEOR measures the accuracy of machine-generated text by analyzing it word-by-word.
- ROGUE measures the accuracy of machine-generated text by analyzing substrings, which are groups of words inside of a sentence.
- CIDEr measures accuracy by machine-generated text to possible options provided by human volunteers.

Each of these metrics is scored on a scale from 0 to 1, with 0 being no matching text to 1 being identical text.

- An average score is around 0.5.
- A good score is around 0.6 to 0.7.
- An excellent score is around 0.8 to 0.9.

A 1 is impossible because that would be exactly the same as human writing, which is not possible with current machine learning methods.

To train artificial intelligence to make captions, you have to train it with a dataset. Two of the most popular are:

- MSCOCO, which has almost 300,000 images and five captions to go along with each picture. It also has categories for different types of pictures with different types of objects to allow for a variety of training data. MSCOCO is also the industry standard.
- Flickr30k, which has 30,000 images and 158,000 captions. Each image has multiple captions so the description can be more detailed. Each picture also has a “bounding box” that makes it easier for the AI to pick out the objects.