CIDEr

CIDEr is used to evaluate image description tasks.

1 TF-IDF vector

The calculation formula of TF-IDF vector is as follows:

$$g_k(s_{ij}) = \sum_{w_l \in \Omega} h_k(s_{ij}) log \frac{|I|}{|I_p : w_l \in I_p|}$$

Among them, Ω is the set of all n-grams, $h_k(s_{ij})$ is the number of times the phrase w_l appears in the reference sentence s_{ij}, $|I|$ is the data The number of all images in the set, $|I_p : w_l \in I_p|$ is the number of images containing the phrase w_l.

2 Definition of CIDEr

The main idea of CIDEr is to treat each sentence as a document, then calculate its n-gram TF-IDF vector, and then use cosine similarity to measure the semantic consistency of the candidate sentence and the reference sentence.
The calculation formula of CIDEr is as follows:

$$CIDEr_n(c_i, S_i) = \frac{1}{m} \sum_j \frac{g_n(c_i) \cdot g_n(s_{ij})}{||g_n(c_i)|| \, ||g_n(s_{ij})||}$$

Among them, c_i is the sentence generated by the model, S_i is the reference sentence set, m is the number of reference sentences, n is the length of n-gram, $g_n(c_i)$ is the length of the sentence generated

by the model TF-IDF vector, $g_n(s_{ij})$ is the TF-IDF vector of the reference sentence.

3 Advantages and Disadvantages of CIDEr

3.1 Advantages of CIDEr

- The correlation between CIDEr's evaluation results and manual evaluation is usually high, so it can be used as a reliable automatic evaluation indicator.

3.2 Disadvantages of CIDEr

- The calculation of CIDEr is relatively complex and may be more time-consuming to calculate.
- The performance of CIDEr may be affected by the corpus used for training. If the corpus is insufficient or unrepresentative, the evaluation results may be inaccurate.

References

https://arxiv.org/pdf/1411.5726.pdf