VQA Accuracy Proposal

Ali Gholami

April 29, 2019

Abstract

This document contains a new accuracy equation.

1 Current Metric (VQA Accuracy)

Assume there are 10 words in the vocabulary. Let's investigate three major cases.

$$\vec{GT} = [5, 3, 2, 0, 0, 0, 0, 0, 0, 0]$$

1.1 Model 1

$$\vec{PV} = [1, 0, 0, 0, 0, 0, 0, 0, 0, 0]$$

1.2 Model 2

$$\vec{PV} = [0, 1, 0, 0, 0, 0, 0, 0, 0, 0]$$

1.3 Model 3

$$\vec{PV} = [0, 2, 0, 0, 0, 0, 0, 0, 0, 0]$$

In all cases, the computed accuracy is 100%. While it may seem plausible (since 5 human annotators have already annotated this word as the answer), there is **answer confidence** missing from this metric.

1. repeition should be considered important for the model, since the **negative log likelihood** is pushing the model toward the direction to predict 5.
2. In model 1, although the answer is correct and the learning is being done flawlessly, the metric does not seem to be plausible to compare the models based on it.
3. In model 2, no emphasis is made on the best answer (although predicted answer is also humanely-acceptable but its not the best one).
4. Model 3 is another case of problem 1.

2 New Metric (NZAD: Non-zeros Averaged Dot Product)

In this section, we propose a new method called **NZAD** which investigates two main components while calculating the accuracy:

1. Answer Generation Validity
2. Answer Validity Confidence

We also believe that this method has an edge to the previous method in terms of computational efficiency (non-zeros processed only, but this totally depends on the way we implement it).

$$NZAD(p,t) = \frac{\frac{2}{NZ(t)} * \frac{<p,t>}{||p||_2+||t||_2} + min\{\frac{AGA}{3}, 1\}}{2} \tag{1}$$

where $NZ(t)$ is the number of non-zero elements in t. AGA is the number of agreeing human answers with the predicted answer. Computing the accuracy using 1 for the given model predictions:

- Model 1: 56%
- Model 2: 46%
- Model 3: 49%