Supplementary material of Can Pre-training help VQA with Lexical Variations?

Shailza Jolly
TU Kaiserslautern, Germany
DFKI GmbH, Germany
shailza.jolly@dfki.de

Shubham Kapoor[1]
Amazon Research, Germany
kapooshu@amazon.com

1 Question Length Analysis on VQA2.0 and VQA-Rephrasing dataset

As described in Section 4 of the paper, we computed question lengths for samples in training data of VQA2.0, validation data of VQA2.0, and VQA-Rephrasings. Fig. 1 presents question length distribution for all three subsets. It can be seen that the data distribution of VQA2.0-train is similar to VQA2.0-val as compared to the distribution of VQA-Rephrasings. Therefore, current VQA models perform well for samples drawn from VQA2.0-val and fail to perform well on rephrasings split of VQA-Rephrasings.

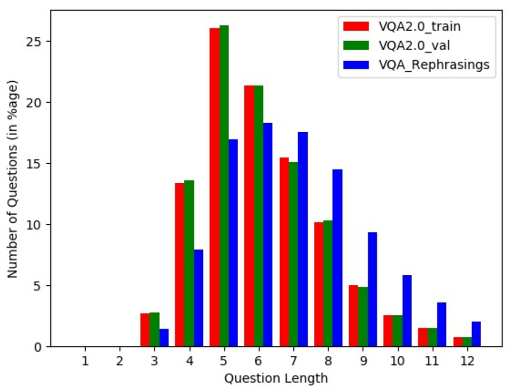

Figure 1: Dataset statistics about the number of questions (in percentage) with varying lengths for three subsets of VQA namely training and validation data of VQA2.0, and VQA-Rephrasings.

2 Comparison of full question and only keywords from question embeddings.

As described in Section 4.3 of the paper, we compared SBERT and GRU embeddings for $S1$ with $S2$ using cosine similarity. $S1$ is a question from VQA-Rephrasing dataset and $S2$ is an ordered sequence of keywords obtained from $S1$. Fig 2 shows the distribution of number of samples in similarity range defined on x-axis. It can be clearly seen that SBERT embeddings are more in higher ranges of cosine similarity as compared to GRU embeddings. Therefore, it can be concluded that pre-trained language encoders (SBERT) latch on keywords.

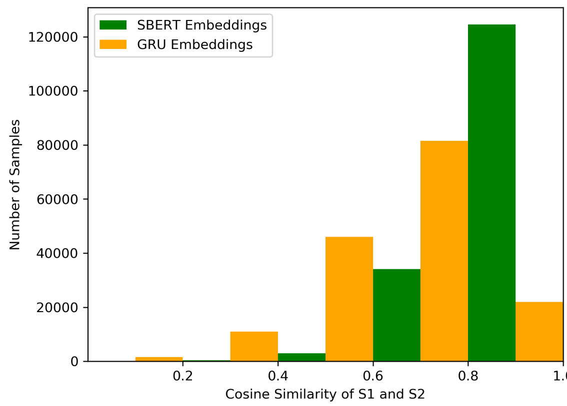

Figure 2: Distribution of cosine similarity of sentence $S1$ and $S2$. $S1$ is a question from VQA-Rephrasing dataset and $S2$ is an ordered sequence of keywords obtained from $S1$.

[1] The work was done prior to joining Amazon.