Question Encoding for Robust Q/A Networks

Arnav Vaid, Anaga Rajan, Chris Correa

Department of Electrical Engineering and Computer Sciences, University of California at Berkeley

Abstract

There are many ways to syntactically ask the same question. Can we train a machine learning model to map similar questions to a single syntax-invariant representation? We propose using neural techniques to learn a mapping from a typical word-embedding question representation to a syntax-invariant one. To evaluate its utility, we attempt to train a question-answering neural model, incorporating our new embeddings.

Motivation and Related Theory

Much progress has been made in learning semantically meaningful representations of words in continuous vector spaces. In this spirit, efforts are being made in learning similar representations for sentences. These embeddings can be learning in an unsupervised manner (techniques like SkipThought [3]), while others can be learned through training for a specific task.

We propose to learn these embeddings by training on sentence similarity tasks. Using a Self-Attentive RNN model [1], we construct an output sequence of new embeddings represented as matrix *M*. We propose an addition to this network which condenses variable length embedding sequences to a single vector representation, allowing explicit comparisons between two questions. We hypothesize that this intermediate embedding sequence *M* produced by the model will result in a more syntactically invariant representation of a question.

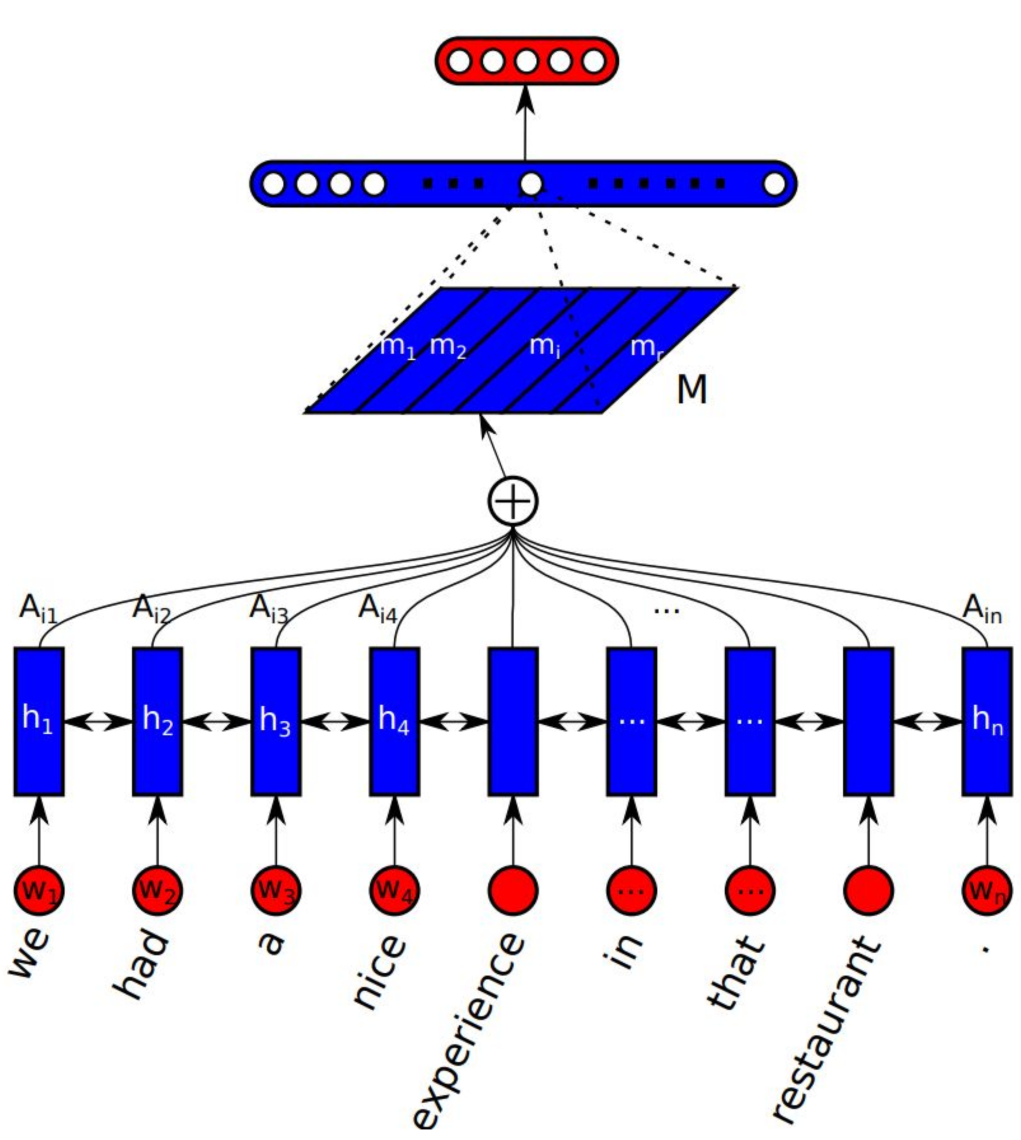


Figure 1. Question Encoding Model Architecture, which produces the *M* matrix of sentence embeddings.

This representation may improve the performance of a question-answering model. QA networks [2] are tasked with finding the answer to a given question within an article. We hypothesize that a QA network might converge faster or achieve more robust performance, if trained on this new learned question representation.

Methodology and Experiments

We trained the Question Encoding (QE) network using the Quora Question Pairs dataset, which contains question pairs with a label indicating similarity. We create the *M* Matrix the same way as [1]. A standard BiLSTM is fed in GloVe word vectors and generates hidden states which are consolidated into the *H* matrix. A self-attention matrix is computed using a feed-forward network, and then multiplied with *H* to get the new embeddings, or the columns of the *M* matrix.

$$H = (h_1, h_2, \dots, h_n)$$

$$A = softmax(W_{s2} tanh(W_{s1} H^T)$$

$$M = AH$$

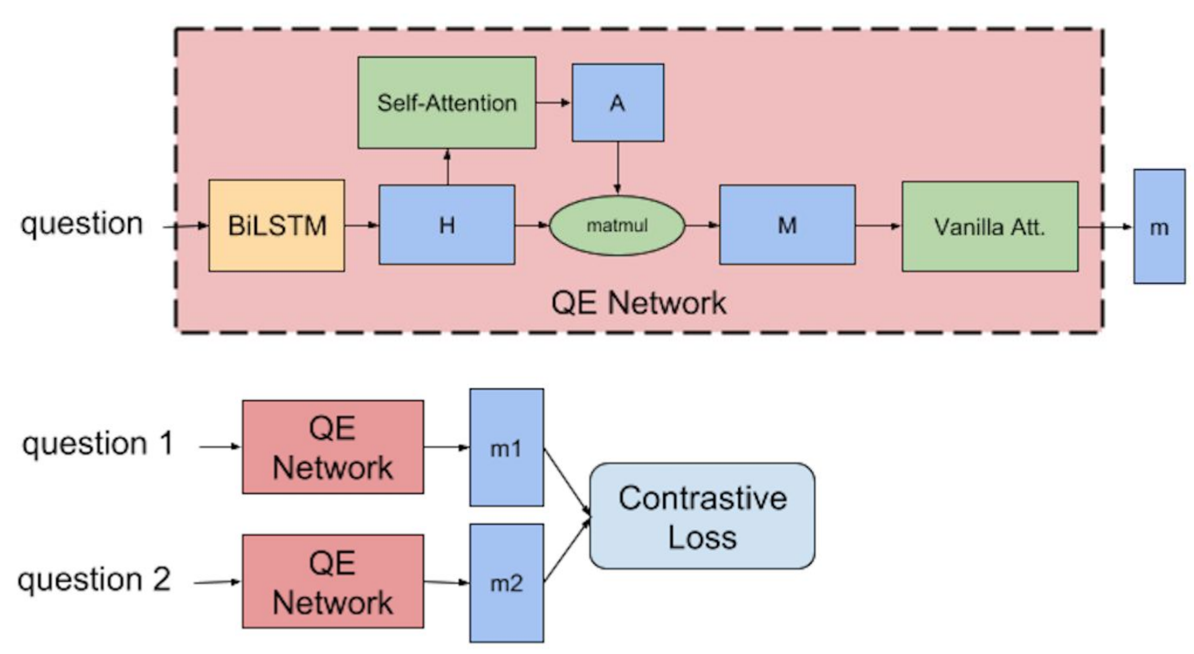


Figure 2. Self attention over the *M* matrices to produce summary vectors m.

Contrastive Loss:
Though we pass in the matrix *M* as our sentence embedding into our DCN network, we apply a vanilla attention module to generate a "summary vector", m, that can be used to calculate loss during training. We opted for a contrastive loss function that is more suited towards calculating embedding similarity rather than standard softmax that optimizes for classification.

$$D_W(X_1, X-2) = ||X_1 - X_2||^2$$

$$L(X_1, X_2) = (1-Y)\frac{1}{2}(D_W)^2 + (Y)\frac{1}{2} max(0, m - D_W)^2$$

DCN:
We pass the Question Encoding Network embeddings as inputs into the DCN network [2], a deep QA neural network. Metrics used are raw accuracy and F1 scores.

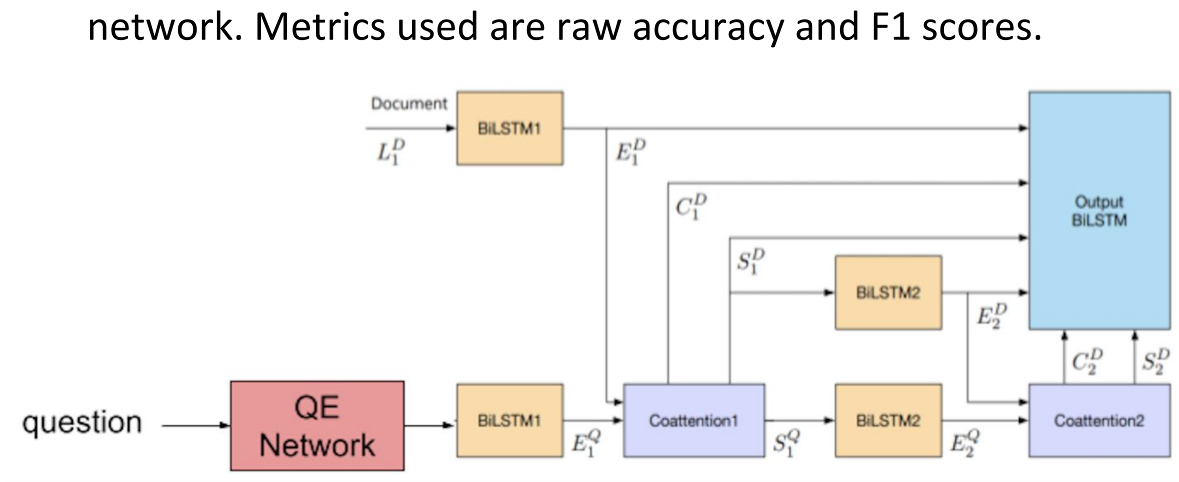
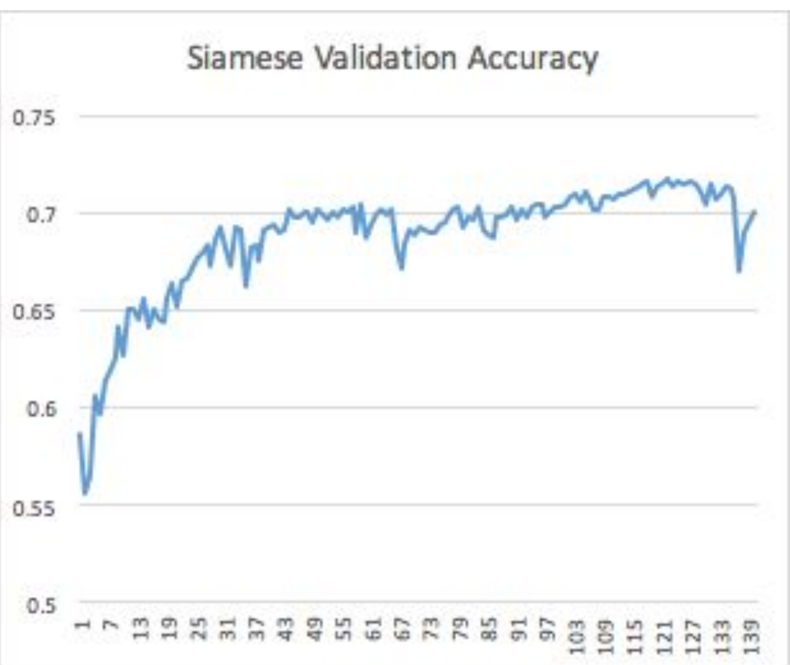


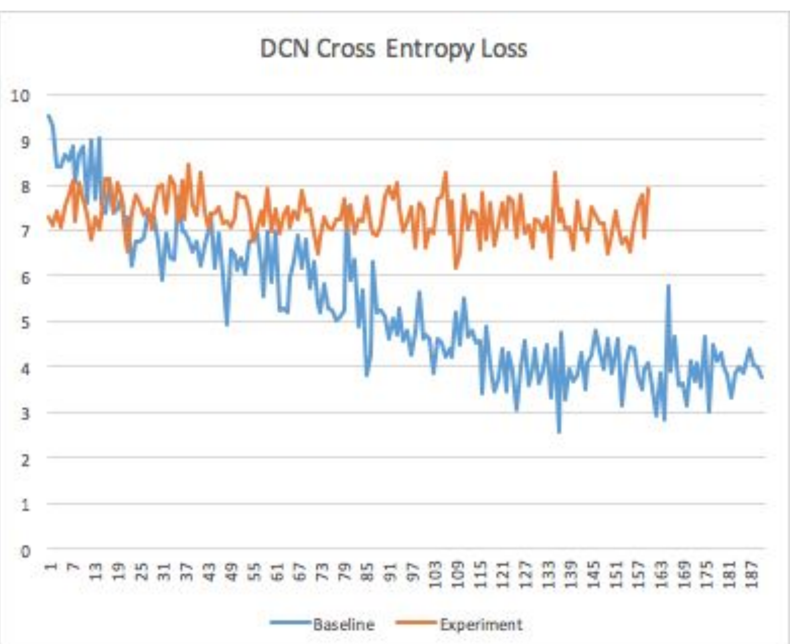
Figure 3. Model Architecture. The columns of *M* are fed into the DCN instead of the word embeddings.

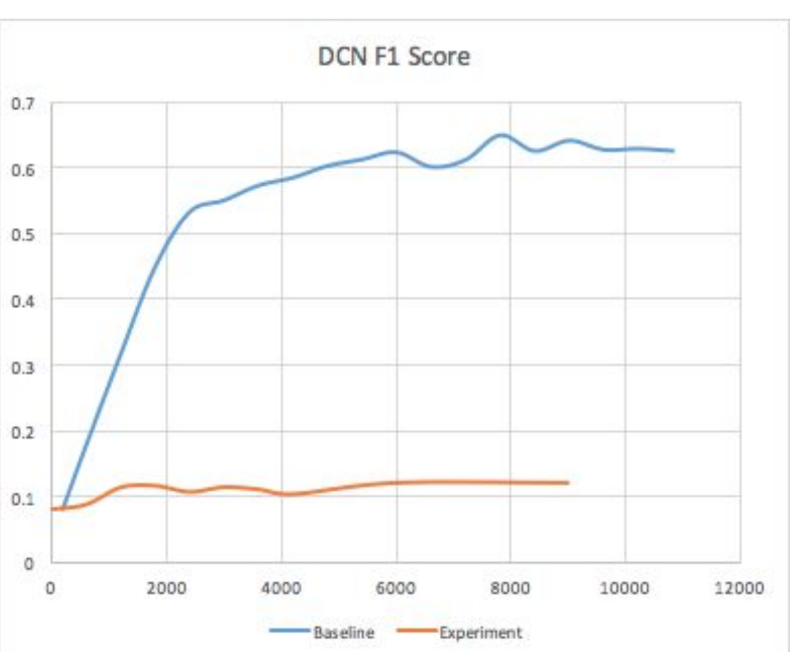
End Goals

In this experiment we hope to improve on the current, state-of-the-art Dynamic Coattention Network (DCN) model for question answering tasks. As per our hypothesis, we believe that incorporating the sentence-similarity embedding to the input of our DCN will improve its performance by bolstering its ability to ignore semantic variance.

Results







We found that simply training the DCN on just the sentence embeddings led to the model performing very poorly. We identify two potential reasons for this. We believe the learned sentence embeddings are well suited for characterizing the sentence into a summery, but individually do not retain the word relationships that GloVe encodes. Second, we believe of the second attention module, the columns of *M* of similar questions don't need to be in the same order to be classified as the same, but should be in the same order for the question answering network. For these reasons we identified potential approaches in the next section.

Future Work

We plan to modify the network to pass the summary vector m directly into the DCN, as one more element in the original sentence embedding. We hypothesize this will improve performance because it will give the DCN both the original embeddings and the syntax-invariant summary vector. This way the DCN will still have access to the GloVe embeddings, which contain specific word relationships, and the summary vector, which contains the meaning of the question, without the orderings of each individual word.

References

[1] Zhouhan Lin, Minwei Feng, Cicero Nogueira dos Santos, Mo Yu, Bing Xiang, Bowen Zhou, and Yoshua Bengio. A structured self-attentive sentence embedding. arXiv preprint arXiv:1703.03130, 2017

[2] Caiming Xiong, Victor Zhong, and Richard Socher. 2016. Dynamic coattention networks for question answering. arXiv preprint arXiv:1611.01604.

[3] Ryan Kiros, Yukun Zhu, Ruslan R Salakhutdinov, Richard Zemel, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. Skip-thought vectors. In NIPS, 2015.