Supplementary material for QUAREL: A Dataset and Baseline Models for Connecting Language with Qualitative Reasoning

Oyvind Tafjord, Peter Clark, Matt Gardner, Wen-tau Yih, Ashish Sabharwal
Allen Institute for AI, Seattle, WA
{oyvindt,peterc,mattg,scottyih,ashishs}@allenai.org

This document provides supplementary material for the paper "QUAREL: A Dataset and Baseline Models for Connecting Language with Qualitative Reasoning."

1 Qualitative Relations

Here is a list of the 19 properties and their relations as used in the QUAREL dataset.

% Scenarios with varying friction
q-(friction, speed).
q-(friction, distance).
q-(friction, smoothness).
q+(friction, heat).
q+(speed, distance).
q+(speed, smoothness).
q-(speed, heat).
q+(distance, smoothness).
q-(distance, heat).
q-(smoothness, heat).
% Scenarios with varying distance
q-(distance, loudness).
q-(distance, brightness).
q-(distance, apparentSize).
q+(loudness, brightness).
q+(loudness, apparentSize).
q+(brightness, apparentSize).
% Speed/time/distance with one kept constant
q-(speed, time).
q+(time, distance).
q+(speed, distance).
% Miscellaneous pairs of properties
q-(weight, acceleration).
q+(strength, distance).
q+(strength, thickness).
q+(mass, gravity).
q-(flexibility, breakability).
q+(exerciseIntensity, amountSweat).

We assume that given a pair of properties, their relation is unambiguous. For richer scenarios, where assumptions might differ, an additional argument specifying the scenario might be required.

Dataset	F1	F1 partial
QUAREL	77.5	84.7
QUAREL^F	85.5	92.9

Table 1: F1 scores for entity tagging of worlds in QUAREL and QUAREL^F.

2 Model details

The following section give implementation details for some of the models in the paper.

2.1 QUASP

The QUASP model is a variant of the neural type-constrained semantic parser in Krishnamurthy, Dasigi, and Gardner (2017). It uses a uni-directional LSTM with hidden state size of 50, taking as input the 1124 dimensional token embeddings. The parser action embeddings are also size 50, while the attention function is a simple dot product. Dropout is 0.2 except for the ELMo embeddings where 0.5 is used. The model was trained using the Adam optimizer (Kingma and Ba 2014) with default settings ($lr = 0.001$, $\beta = 0.9$, $\beta_2 = 0.999$).

2.2 BILSTM

The BILSTM model is similar to QUASP, except we do a separate hyper parameter search to arrive at a two-layer bi-directional LSTM in the encoder. We concatenate the first and last encoder output and directly predict the two-way answer using a linear layer.

2.3 QUASP+

As part of the QUASP+ model we train a separate BIO entity tagger for "world" entities. This tagger uses a bi-directional LSTM encoder with hidden state size of 50 which feeds into a CRF to predict consistent assignments of B-World, I-World, and O tags. The input is 100-dimensional Glove vectors (Pennington, Socher, and Manning 2014).

The tagger obtains good F1 scores, especially when we allow partial overlaps which could still be useful for the downstream application. See table 1 for detailed scores on both QUAREL and QUAREL^F.

As described in the paper, the resulting world tags are used to replace certain phrases in the questions. Otherwise the model is the same as QUASP.

2.4 QUASP+ZERO

For the zero-shot model QUASP+ZERO, we treat the properties as dynamic entities for which we calculate a similarity score to each question token as described in the paper. These scores are then combined with the question attention to predict the property slots in the parse. Otherwise the model is the same as QUASP+. In particular we don't change the input to the encoder, so we don't introduce the "entity embeddings" used in Krishnamurthy, Dasigi, and Gardner (2017).

References

Kingma, D. P., and Ba, J. 2014. Adam: A method for stochastic optimization. *CoRR* abs/1412.6980.

Krishnamurthy, J.; Dasigi, P.; and Gardner, M. 2017. Neural semantic parsing with type constraints for semi-structured tables. In *EMNLP'17*.

Pennington, J.; Socher, R.; and Manning, C. D. 2014. GloVe: Global vectors for word representation. In *EMNLP'14*.