A Appendices

A.1 BertQA Baseline

Our BertQA baseline follows that proposed by Devlin et al. (2019) for the Stanford Question Answering Dataset (SQuAD) (Rajpurkar et al., 2016). Due to the differences in context between ShARC and SQuAD, we augment the input to the BERTQA model in a manner similar to Section 3.1. The distinction here is that we additionally add the decision types “yes”, “no”, and “irrelevant” as parts of the input such that the problem is fully solvable via span extraction. Similar to Section 3.1, let U denote the BERT encoding of the length-n input sequence. The BERTQA model predicts a start score s and an end score e.

$$s = \mathrm{softmax}(UW_s + b_s) \in \mathbb{R}^n \tag{28}$$

$$e = \mathrm{softmax}(UW_e + b_e) \in \mathbb{R}^n \tag{29}$$

We take the answer as the span (i, j) that gives the highest score $s_i e_j$ such that $j >= i$. Because we augment the input with decision labels, the model can be fully supervised via extraction endpoints.

A.2 Creating noisy supervision for span extraction via span matching

The ShARC dataset is constructed from full dialogue trees in which annotators exhaustively annotate yes/no branches of follow-up questions. Consequently, each rule required to answer the initial user question forms a follow-up question in the full dialogue tree. In order to identify rule spans in the document, we first reconstruct the dialogue trees for all training examples in ShARC. For each document, we trim each follow-up question in its corresponding dialogue tree by removing punctuation and stop words. For each trimmed question, we find the shortest best-match span in the document that has the least edit distance from the trimmed question, which we take as the corresponding rule span. In addition, we extract similarly trimmed bullet points from the document as rule spans. Finally, we deduplicate the rule spans by removing those that are fully covered by a longer rule span. Our resulting set of rule spans are used as noisy supervision for the rule extraction module. This preprocessing code is included with our code release.