NLP Project Proposal

Michael Scheriger

For my NLP project, I would like to attempt to perform a question answering task. While researching this topic, I discovered the AI2 Reasoning Challenge, which provides a corpus of 7,787 multiple choice questions. The AI2 Reasoning Challenge also provides a larger corpus comprised of 14 million science sentences that can be used as a resource to answer the questions. The multiple-choice dataset is split into "easy" and "hard" questions. I believe a reasonable goal is to develop a machine learning algorithm that performs well on the "easy" questions and one that performs better that random guessing on the "hard" questions. The paper provides a description of the problem as well as multiple solutions that can serve as a baseline. I will need to do further research in order to become more familiar with state-of-the-art methods for question-answering tasks, but hopefully I can start with a simple SVM or neural network and use it as a baseline.

Paper: http://ai2-website.s3.amazonaws.com/publications/AI2ReasoningChallenge2018.pdf

Data: https://allenai.org/data/arc