Independent Study Report

Vedant Puri

May 2019

1 Introduction

The task assigned to me was to perform supervised learning on scientific procedural texts. This is also known as shallow semantic parsing and in simple terms, my task was to implement code that learnt the label between a predicate and an argument in scientific texts from the training data and "successfully" and competitively predicted a label for a predicate argument pair in the testing/dev set.

2 Preparation

Prior to this study I was just accustomed to using sklearn for Machine Learning as was taught to me in CS589. But as suggested by my advisor it would be a great learning experience to dive into PyTorch and handle this task by implementing code using PyTorch. Consequently it took me a while to understand the library and eventually get used to how it works. I went over tutorials and implemented the following before I got to the actual task at hand:

1. An image classifier that was mentioned on the tutorials that used a multilayer neural network to classify an image in one of many classes. It uses cross-entropy loss as the loss function and runs using stochastic gradient descent.

2. A text classifier that used a bag of words representation as its features to learn and then predict the language that the text was in. There were 3 classes for this task: English, Spanish and French. It was essentially a Logistic Regression classifier and hence used a Negative Log Likelihood loss and ran using stochastic gradient descent.

3 Data

Helpfully enough, my mentor during this project, Sheshera Mysore, helped with the splitting of data into train, development and testing sets. The data was in the form of annotated sentences that can also been seen in a GUI at (ADD MIT LINK HERE).

4   Pre-Processing

- obtaining relevant sections from the annotated dicts - forming predicate, arg, role pairs and storing preds and args as separate lists and the reasoning for doing that -

5   Training

6   Evaluation

- sklearn metrics used

7   Conclusion