PNet - ParseNet

Progress Report

Sai Krishna Karanam karanam.s@husky.neu.edu
Nischal Mahaveer Chand mahaveerchand.n@husky.neu.edu
Varun Sundar Rabindranath rabindranath.v@husky.neu.edu

I. Changes

We have also changed the datasets used for our task. As mentioned in the Edinburd paper (cite), the Hearthstone and Djanjo datasets, are not generalizable and would not be the right data to use for the task (Discussed more elaborately in Section X). As a result, we decided to scrape our own data from the repositories mentioned in the Edinburd paper.

II. Project Plan

Due to the important task of gathering our own data for the task, the data creation and processing phase of our original plan has taken up significant time, resulting in the following new plan.

1) Phase 1: Literature Survey [Done]
2) Phase 2: Data Preprocessing [In-progress]
3) Phase 3: Model training [Future work]

III. Data

A. Problems with existing datasets

HS and DJ are bad. Placeholder

B. Data gathering

Scrape from web.

C. Data preprocessing

processing...

IV. Methods

The two methods used:

1) Base paper placeholder
2) Vanilla seq2seq: To further illustrate the problems with HS and DJ data, we train a vanilla seq2seq model in TensorFlow. Google separately provides a high-level API for the same at google/seq2seq (cite github seq2seq), under the Apache 2.0 license. We run the same model without modification.
As TensorFlow provides the ability to see the training process via TensorBoard, we stopped model training when we saw no visible improvement on the dev corpus. The model trained was a BiDirectional LSTM model, with attention as described in (cite).

V. Experiments

DO you even experiment?

VI. Results

End of humanity is near.

VII. Future Word

DO IN FUTURE!

References

[1] M. Rabinovich, M. Stern, and D. Klein, "Abstract syntax networks for code generation and semantic parsing," arXiv preprint arXiv:1704.07535, 2017.