Sequence to Sequence in SPARQL(Warm-Up)

Anand Panchbhai—Computer Science and Engineering IIT Bhilai, India, anandp@iitbhilai.ac.in

I. INTRODUCTION

This document contains the information regarding the warm up task for the project **A Neural QA Model for DBpedia** done by the me and corresponding inference made in some of the cases.

There are 4 Sections:

1) Initial Impressions
2) Paper Survey
3) Download edit a sample template and train a Neural SPARQL Machine model on a DBpedia class.
4) Read the blog of last year's student to understand what has been achieved so far and reproduce the experiments.

II. INITIAL IMPRESSIONS

A. Issues faced and corresponding fixes

The repository had a few issues that i tried to fix by creating a pull request. As i worked, there were more subtle mistakes in the repository that i will try to correct in the coming months. The pull request made by me can be found Here, the request was accepted.

While using the nmt submodule I came across a lot of errors, to overcome them I cloned the nmt repository separately and learnt how to use it. The results i obtained on doing so are mentioned in detail in the coming sections.

III. PAPER SURVEY

From the initial survey i came across the following points that might help determine the direction of the project:

1) Neural Machine Translation for Query Construction and Composition (2018).
   a) Camparision with state of art in QALD is to be done.
   b) The manual curation of the template (24 in total in the paper), poses a big restriction on the applicability of the model in real world environment.
2) SPARQL as a Foreign Language (2017)
   a) More metrics to measure the performance gives a wider view of the usability and overall applicability of the model. The current model should incooperate the use of F-score to get a well rounded information of the model we intend to built.
   b) Attention method has surpassed LSTMs and RNNs in various fronts, it will be worth while to ponder in the direction to use attention in our model.
   c) Generating domain independent templates is a future scope of this paper and should be one of the goals of the proposal.
   d) Determining and the use of correct number of variables in the queries poses another challenge that need to be addressed in future works.

IV. REPRODUCE THE EXPERIMENTS: NSPM

On reproducing the experiment using the monuments and testing the results the performance recorded are as folows:

Monument	perplexity	BLEU
dev	1.26	87.3
test	1.04	96.6

Note : *The whole model was divided into 3 sets: train, dev and test. The table above depicts the results of the dev and the test set after 12000 iterations. Runtime log can be found at Click here.*

After following the Wiki link : https://github.com/AKSW/NSpM/wiki/DBpedia's-GSoC-warmup-tasks, I was able to reproduce the results by creating the templates as asked for the ontology class of my choice, here:person. The results recorded are as follows:

Person	perplexity	BLEU
dev	1.27	86.4
test	1.04	96.4

Note : *The whole model was divided into 3 sets: train, dev and test. The table above depects the results of the dev and the test set after 12000 iterations. Runtime log can be found at Click here.*

V. DOWNLOAD EDIT A SAMPLE TEMPLATE AND TRAIN A NEURAL SPARQL MACHINE MODEL ON A DBPEDIA CLASS

Reproduction of Aman's pipeline gave me the following results: Click here., the instructions seemed be a bit vague and step 6 used parts of the files that were not created by in the the earlier steps. The sixth step is given here for reference.

```
python final_formatting.py data/GS-v3.csv
data/annotations_place_v2.csv
```

The resulting annotations file had 707 templates which was not feasible for me to run on my laptop. Thus I decreased the number of rows and ran the NSPM model on the data from the above mentioned pipeline.