

Evaluating and Developing Methods of Generating Code-Switched Data 07-400, Spring 2022

Abhishek Vijayakumar,
under Prof. Alan W. Black, Language Technologies Institute
<https://inkyubeytor.github.io/category/code-switching.html>

February 14, 2022

1 Major Changes

The primary goal of this project has not changed since the previous milestone. An additional secondary goal is to train a transliteration model following modern practices for Hindi romanization in colloquial settings.

2 What You Have Accomplished Since Your Last Meeting

I have finetuned BERT models on data generated from a variety of methods with two sets of pretraining options. I have run GLUECoS evaluations on these models and collated results. I have also found several issues in the existing benchmark (see Surprises).

3 Meeting Your Milestone

The original Feb 14th milestone was to implement 4 distributional methods and run evaluations. I have completed this milestone. The original March 1st milestone was to implement the GLUECoS evaluation locally. While I can run these evaluations now, we are planning to change the structure of these evaluations based on a surprise noticed during the process of testing this local evaluation (see Surprises).

4 Surprises

The local GLUECoS evaluation has multiple possible flaws in the evaluation. The dataset used for the NLI task is an outdated version with several issues in the cleaned text which may affect evaluation. I am planning to find the new version of this data or clean it myself.

The dataset used for the QA task produces extremely high variability in results depending on what fold of the data is used as validation data, likely due to improperly shuffled data or a low amount of overall data. I am planning to implement nested cross validation for training and testing on all tasks to fix this problem.

5 Looking Ahead

During the next two weeks, I plan to acquire the new dataset and implement more rigorous validation and testing practices across all the working tasks. I plan to re-evaluate the baseline model used by the creators of the benchmark with the more rigorous evaluation and then evaluate several fine-tuned models on the same version.

6 Revisions to Your Future Milestones

I am replacing my next milestone with the goal of improving the robustness of the evaluation. I am also likely to move up my milestone of implementing additional generative methods and start that process soon, because the amount of time taken for nested cross validation will likely prevent me from making much progress.

7 Resources Needed

No further resources are needed for this project at this time. I have obtained a new Hinglish dataset from my advisor.