

[urdu]rmAmiri

Assignment 2: Filtering

Improving Source-to-Target Sense Projection (English–Urdu)

Chirooth Girigowda
girigowd@ualberta.ca
University of Alberta

Sahir Momin
smomin1@ualberta.ca
University of Alberta

1 Introduction

In this assignment we implemented ExpandNet on Urdu. We performed translation with an Urdu pre-train Language model "alif-urdu-llm". We then used dbalign along with the dictionary that we found via wikiextractor and other dictionary resources(cite all of them), with this combined dictionary and translations we get the alignment. Finally, we perform project senses of the english tokens onto the urdu words from the alignment. (if space permits include information about the data and code we recieved - can include this below) We project the BabelNet senses, given to us along with it's gloss in *se_{gloss}.tsv*

2 Baseline

2.1 Implementation of the LLM Baseline

2.2 LLM Used and Access Method

2.3 Prompt Design and Justification

2.4 Number of Senses Generated per Synset

2.5 Examples, Errors, and Notable Cases

3 Method

3.1 Translation

3.2 Alignment

3.3 Filtering

4 Analysis

4.1 Correct vs Incorrect Projections

4.2 Contextually Possible vs Impossible Equivalents

4.3 Sources of Error

4.4 Comparison with LLM Baseline

4.5 Automatic Evaluation

4.6 Impression

5 Conclusion