

Report - Homework #5

1 Prompting Large Language Models

For the last part of this project, I implemented my own example selection method, inspired by the idea of increasing the diversity of the demonstrations. Suppose we are going to pick k examples for the prompt. Based on the embedding similarity method, instead of selecting the top- k examples at once, we select the top example one at a time. After each example is selected, we first delete it, subtract its embedding from all other remaining examples and recalculate the similarity score. In this way, we pull all other examples away from the direction of the selected example each time, which should increase overall diversity.

In my experiment, uniform random sampling got 37.5% accuracy, embedding-based similarity search got 70% while my own method got 67.5%. Both embedding-based and my own method make significant progress over uniform random selection.

For example utterance "what river runs through m0", the output of each method is as follows.

- Uniform random sampling: `answer (intersection (river, m0))`
- Embedding-based similarity search: `answer (traverse_1 (intersection (river , loc_2 (m0))))`
- My own method: `answer (traverse_2 (intersection (river , loc_2 (m0)))))`

We can see that both embedding-based and my own method predicted more accurate semantic parsing results.