Project Proposal: Semantic Parsing of Propositional Logic in McGill University's Course Pages using Large Language Models

Liam Scalzulli, Jeff Zhang, Sam Zhang

2023-10-27

Background In the domain of Natural Language Processing (NLP), large language models (LLMs) have shown promise in various tasks, from text generation to question-answering. However, the capacity of LLMs to interpret and translate natural language into formal logical constructs remains an area for exploration, with minimal prior work. One prevalent real-world example of such a challenge exists in academic course descriptions, where prerequisite and co-requisite information often melds structured logic with freeform language.

Objective In McGill University's course pages, course requisites use a semi-formal structure, as they contain keywords such as `and` and `or` and look rule-based (Figure 1), but do not actually follow any pre-determined formal structure and are rather natural-language based, leading to some courses' pre-requisite being non-parsable using a deterministic parser. (Figure 2)

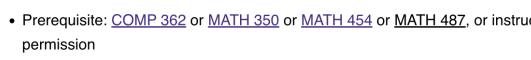

Figure 1: Course Prerequisites for COMP553

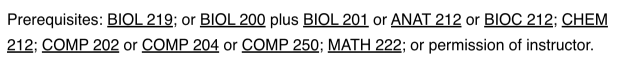

Figure 2: Course Prerequisites for BIOL395

We aim to generate a rule-based, tree-like output in JSON format, e.g. for the string (COMP202 or COMP208) and COMP250, the desired output could be

```
["&",["|",["COMP202"],["COMP208"]], ["COMP250"]]
```

or

```
{ "prerequisites": { "operator": "AND", "children":
[ { "operator": "OR", "children": [ "COMP202", "COMP208" ] }, "COMP250" ] }}
```

Problem Statement Given the semi-structured nature of course descriptions, which incorporate logical constructs expressed through natural language (e.g., using terms like "AND" and "OR"),

can LLMs be fine-tuned to effectively parse and translate these descriptions into clear propositional logic representations?

Methodology A rough planning of the project's methodology is as below:

Course requisite strings will be scraped from McGill University's eCalendar and a portion of the data will be manually selected and labelled for training and testing.

We will use a mainstream method to fine-tune an LLM (e.g. fine-tuning a GPT via OpenAI's API, fine-tuning a pre-trained Llama2, etc.) using the samples above.

Experiments with the fine-tuned model will be driven against other mainstream LLMs (GPT, Llama, PaLM, Wenxin Yiyan etc.), where their performances will by evaluated by pre-existing or self-made metrics.

Significance This project can be significant in the realm of NLP in a few ways.

First, it may provide insights into the capabilities and limitations of LLMs in bridging natural language with formal logic, potentially filling gaps in existing research.

This paper may also offer methodologies or inspirations that may be generalizable to other domains requiring a semantic understanding of logical constructs, such as in a more complex, conversational, or textual context.

Finally, we are also happy to contribute to the growing body of open-sourced knowledge on fine-tuning and domain adaptation of LLMs for specific tasks.