Formalizing Natural Language: Cultivating LLM Translations Using Automated Theorem Proving

Richard Thompson[1[0009−0001−6541−1092]], Adam Pease[1[1111−2222−3333−4444]], Roberto Milanese Jr.[1[0009−0009−5107−162X]], Jarrad Singley[1[1111−2222−3333−4444]], and Angelos Toutsios[1[0009−0009−6064−5154]]

Naval Post Graduate School, 1 University Circle, Monterey, CA 93943, USA

1 Introduction

In recent years Large Language Models (LLMs) have improved significantly in the ability to ingest massive amounts of information and produce statistically relevant answers to queries. Unfortunately there is a lack of formalism with this approach, and real reasoning is absent in LLM responses. Symbolic AI on the other hand is able to logically reason through a problem with concrete steps, but draw from a relatively small quantity of formal axioms, and are thus limited to specific domains. The neuro-symbolic approach draws from the benefits of both, and limits the disadvantages inherent in each. We present a system that can take natural language and translate sentences into higher order formal axioms using a LLM.[1] Automated theorem proving (ATP) is then used to ensure our growing knowledge base is consistent, and to answer questions from our knowledge base in a provable way.

Human language is expressive and diverse, more so than can be expressed in First Order Logic [4]. Thus to create our system, we use a large ontology called Suggested Upper Merged Ontology (SUMO) [2], which contains over 230,000 axioms formalized with expressive higher order logics. Axioms in SUMO are continuously being checked for consistency using the Vampire Eprover [1]. Our system is constantly improving, being able to successfully translate growing categories of sentence structures.

2 System Overview

The dataflow of our system is shown in Figure1. The target user of our system is a domain expert composing policy. The user will input a policy document, written in natural English, to a computer. The text is first sent to a Metaphor Handler, which uses an LLM to detect and replace metaphorical words and phrases. The result is sent to a Sentence Simplification module, a combination of heuristics and an LLM which rephrases the sentences into simpler grammatical structures,

[1] Code repository for our system is found on Github at: https://github.com/ontologyportal/sumonlp

similar to the sentences on which our Language to Logic (L2L) translator was trained.

Every word in the WordNet lexicon [3] is mapped to SUMO, however even with such an expansive dataset, there are words and organizations that will be encountered that are unknown. Rather than fail, the system will attempt to produce logic using Stanford's Stanza Name Entity Recognizer (NER) to detect the word type. The word is replaced with an appropriate label in the sentence, which our L2L translator has been trained to handle. After translation to logic is complete, the label is replaced with the original word. Thus, successful translation is still possible even when an unknown word is encountered.

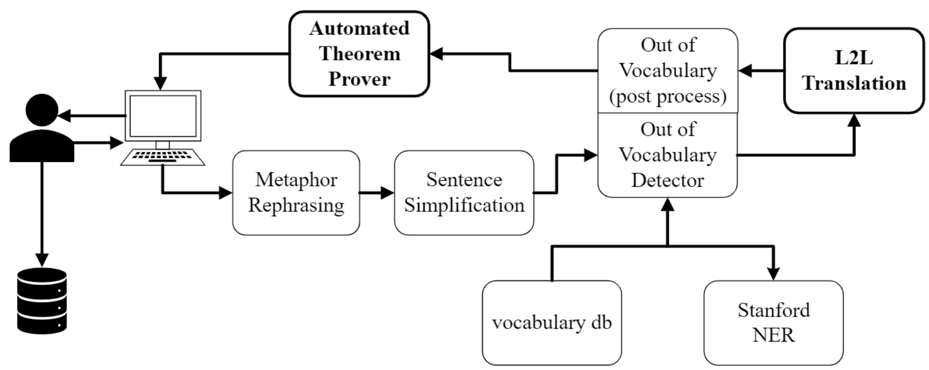

Fig. 1. Language to logic architecture

2.1 Language to Logic Translation and Theorem Proving

The L2L translator is trained from a synthetic corpus of around 12 million sentences with their corresponding logic translations. Language logic pairs are built by creating a frame structure for each sentence and filling in the slots of the frame from the contents of SUMO; the objects, processes and relationships. Because we understand the intended semantics of the frame, we are able to generate equivalent english and logic expression of the frame. This is in contrast to methods that otherwise rely on untrained, non-learning methods, or human manual generation of a training set.

After translation to logic, the new axioms are added to the SUMO ontology. The Vampire ATP is then tasked with finding and contradictions that might have been introduced. The user can then use SUMO to ask questions of the knowledge base. These questions are similarly translated to a formal conjecture, which Vampire is used to respond to.

References

1. Kovács, L., Voronkov, A.: First-order theorem proving and Vampire. In: Sharygina, N., Veith, H. (eds.) CAV. LNCS, vol. 8044, pp. 1–35. Springer (2013)
2. Niles, I., Pease, A.: Towards a Standard Upper Ontology. In: Proceedings of the International Conference on Formal Ontology in Information Systems-Volume 2001. pp. 2–9 (2001)
3. Niles, I., Pease, A.: Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. In: Proceedings of the IEEE International Conference on Information and Knowledge Engineering. pp. 412–416 (2003)
4. Pease, A.: Choosing a Logic to Represent the Semantics of Natural Language. In: Logic and Argumentation: 4th International Conference, CLAR 2021, Hangzhou, China, October 20–22, 2021, Proceedings 4. pp. 504–512. Springer (2021)