

SUMO's new bAbI: A Neuro-Symbolic Approach to Natural Language Inference

Adam Pease

1 Introduction

We are conducting experiments in machine translation from natural language to logic using SUMO <https://github.com/ontologyportal/sumo> terms. The code to create the datasets for training is in `com.articulate.nlp.GenSimpTestData` at <https://github.com/ontologyportal/sigmanlp>.

The ML training resource used is <https://github.com/tensorflow/t5>

2 Language Generation

This SigmaNLP code generates language-logic pairs designed for training a machine learning system. Several approaches are used:

- instantiate relations with arguments of appropriate types and then generate NL paraphrases from them
- run through all formulas in SUMO and generate NL paraphrases
- build up sentences and logic expressions compositionally

The compositional generation is potentially the most comprehensive. It consists of building ever more complex statements that wrap or extend simpler statements. Originally, this meant starting with a simple subject-verb-object but now encompasses a host of different constructs:

- six tenses (part, present, future and progressive or not) and correspond to SUMO expressions relative to a diectic 'Now'
- negation (for action sentences and separately for attitude wrappers that require higher-order logic, such as `believes`, `knows` etc)
- frequency-based selection of nouns and verbs
- human names and social roles
- restrict human actions with `CaseRole` of `agent` to be `IntentionalProcess(es)` and otherwise use the `CaseRole` of 'experiencer'
- use WordNet's noun and verb morphological exception lists

- counts of objects (0-10) are used randomly a small portion of the time, divided between using numerals some times and digits other times
- quantities of substances with units are used randomly a small portion of the time
- Some sentences have metric dates and times, in several formats
- Queries ('what','who') are generated sometimes, and with a question mark
- WordNet verb frames are used that index verbs to a context of use. However, this is quite limited compared to VerbNet and FrameNet, and still allows many semantically non-sensical sentences to be generated, while also missing some common sentence forms, especially use of prepositions
- Sentences with all variations of SUMO's modalAttribute are generated (Possibility, Obligation etc)
- Sentences with all variations of SUMO's PropositionalAttitudes are generated believes, knows, says etc
- Sentences with says have their content quoted
- Imperative sentences with the English "You (understood)" form

3 User Interface

The system has a simple top-level command line interface

Valid commands are ask, add, clear or quit

"ask" a question from the knowledge base.
Example: "ask Does CO2 cause global warming?"

"add" will append a new sentence or file to the knowledge base.
Example: "add CO2 causes global warming."
Example: "add climate_facts.txt"

"clear" will completely clear the knowledge base.
Example: "clear"

"last" will show the progression through the pipeline of the last added sentence or file.
Example: "last"

"list" or "kb" will display the knowledge base.
Example: "list"

"prover" or "test" will run the Vampire prover on the knowledge base, searching for contradictions.
Example: "test -t 40" # runs for 40 seconds
Example: "test" # runs for 60 seconds

"mh" will run just the metaphor translation portion of the pipeline.

Example: "mh The car flew past the barn."

"ss" will run just the sentence simplification portion of the pipeline.

Example: "ss He who knows not, knows not he knows not, is a fool, shun him."

"oov" will run just the out of vocabulary handling portion of the pipeline.

Example: "oov Bartholemew used the doohicky as a dinglehopper."

"l2l" will run just the language to logic portion of the pipeline.

Example: "l2l A bird is a subclass of animal."

"hist" or "history" will print the commands run.

Example: "hist"

"quit" will exit the interface.

Example: "quit"