Programming Languages: Final Project Proposal

John Doherty

April 7, 2018

In this class, we have learned about the importance of Context Free Grammars in interpreting programming languages. These Grammars are used in constructing parsers for programming languages that can implement efficient REPLs. An interesting problem, is if such a Grammar can be learned. While this may not have much application in the programming languages field (unless one wishes to infer the evaluator step of the interpreter too) it is important in natural language processing. This is because one can mostly model natural languages using a Context Free Grammar [2]. Thus an algorithm which can infer Context Free Grammars is an important step in inferring grammars generating natural languages.

The problem that is answered in the paper of interest is given a set of strings that exist in a language described by a Context Free Grammar, can that Grammar be inferred? This has been shown to be possible by [1] for a certain subset of CFGs, called NTS (non terminal symbol) Grammars. The paper used their proposed algorithm to win the Omphalos Competition, which was a competition to infer Grammars from positive and negative examples from the language (though his method only requires positive examples).

As a part of my project, I plan to implement the algorithm described in the paper which accepts a set of strings and determines a CFG that can generate the given strings. I will test the algorithm's performance determining CFGs from hand made production rules. Once tested on tailored examples, I will test how it performs on different programming languages, such as Haskell, as well as natural languages such as English. When testing these languages I will look to see if the algorithm is able to identify major grammatical structures, such as "if a then b else c" or "Subject Verb Predicate." Once the project is complete, I will be able to characterize the performance of the algorithm, identifying its strengths and weaknesses

References

[1] CLARK, A. Learning deterministic context free grammars: The omphalos competition. *Machine Learning 66*, 1 (2007), 93–110.

[2] SHIEBER, S. M. Evidence against the context-freeness of natural language. In *Philosophy, Language, and Artificial Intelligence*. Springer, 1985, pp. 79–89.