Syntax Based Natural Language Models for Code Completion of Android API Calls

Matthew Kilgore

Veselin Raychev, Martin Vechev, and Eran Yahav, in their paper “Code Completion with Statistical Language Models,” proposed a natural language model of programs. Their model views each program as a series of sentences, with words corresponding to function calls. However, the model suffers some deficiencies:

- The model is unable to handle arbitrary loops, instead requiring a bound on the number of iterations. This eliminates many possible programs.
- Moreover, the method of constructing executions counts every occurrence of a function call in a loop once for every iteration, potentially skewing results towards occurrences common in loops.

We address this by considering an alternate model, building the sentences from the parse-tree of the program in a manner that attempts to limit these concerns. We also expand the model to consider a MEMM for API call selection.

Code: https://github.com/KleinFourGroup/NLP_PL