PHOG: Probabilistic Model for Code

Pavol Bielik PAVOL.BIELIK@INF.ETHZ.CH
Veselin Raychev VESELIN.RAYCHEV@INF.ETHZ.CH
Martin Vechev MARTIN.VECHEV@INF.ETHZ.CH
Department of Computer Science, ETH Zürich, Switzerland

Learned Probabilistic Higher Order Grammars (PHOGs)

We provide the learned TCOND programs that parametrize PHOG trained on JavaScript code corpus. The programs for non-terminal and terminal symbols are shown in the Table 1 and Table 2 respectively. Because we learn a TCOND program depending on the parent non-terminal type, each table contains multiple programs, one for each parent non-terminal type.

The learned context often includes traversing to one or more positions in the AST that are similar to the position being predicted. For example:

- for keys in a JSON object (i.e., a dictionary) it is the position of the key (denoting how many other keys were already defined) and previously defined key,
- for strings, the learned context includes two previous string values seen in the program,
- for array index, it is the index used previously.

As a concrete example consider the learned program for predicting strings (`LiteralString` in Table 2):

```
PrevNodeType WriteValue PrevNodeType
WriteValue PrevLeaf PrevNodeType WriteValue
```

By looking at the instructions, we can see that the program first conditions on two previous string values seen in the AST by using `PrevNodeType` followed by WriteValue instructions. This is intuitively a good strategy and it allows the model to correctly predict common patterns such as `"[" + attrs + "]"`. where the model would condition the prediction of `"]"` on previous string `"["` (and two other values).

Table 1. Learned conditioning program for PHOG to predict non-terminals (node types).

PARENT NON-TERMINAL SYMBOL	TCOND PROGRAM
root	Empty
Program	Left WriteType Left DownLast WriteType NextLeaf WriteValue
ExpressionStatement	Up WritePos PrevNodeContext DownLast WriteType PrevNodeType NextLeaf WriteValue
CallExpression	PrevDfs WriteType WriteValue PrevDfs WriteValue
MemberExpression	WritePos Up PrevNodeContext DownFirst WriteType
Property	Up PrevNodeContext DownLast WriteType Left DownFirst WriteType Right PrevDfs Right WriteValue
ObjectExpression	WritePos Left WriteValue PrevLeaf WriteValue
VariableDeclaration	PrevDfs WriteType PrevDfs WriteType PrevLeaf WriteValue PrevNodeType PrevLeaf WriteValue
VariableDeclarator	Up WriteValue PrevLeaf WriteValue PrevLeaf WriteValue WritePos
FunctionDeclaration	WritePos PrevDfs WriteValue PrevNodeType WriteValue
BlockStatement	Left WriteType Left PrevDfs Up WriteType
AssignmentExpression	Left WriteType PrevNodeContext Right WriteType PrevLeaf WriteValue PrevLeaf WriteValue
ArrayExpression	Left WriteType PrevNodeContext Right WriteType Up DownLast PrevNodeContext DownFirst WriteType PrevLeaf WriteValue
FunctionExpression	PrevDfs WriteValue PrevDfs WriteValue PrevLeaf WriteValue PrevLeaf WriteValue
ReturnStatement	PrevLeaf Right WriteValue PrevDfs WriteValue PrevLeaf WriteValue PrevLeaf WriteValue
ArrayAccess	PrevDfs WriteType WriteValue PrevNodeType DownFirst WriteValue PrevLeaf WriteValue
IfStatement	Left WriteType Right PrevLeaf WriteValue PrevLeaf PrevNodeType WriteValue PrevLeaf WriteValue
UnaryExpression	Up WriteValue PrevDfs WriteValue PrevLeaf WriteValue PrevLeaf WriteValue
ThrowStatement	Up PrevNodeContext DownFirst WriteType PrevLeaf WriteValue PrevLeaf PrevNodeType WriteValue
NewExpression	PrevDfs PrevNodeContext WriteType DownFirst DownLast Right WriteType PrevLeaf WriteValue PrevLeaf PrevNodeType WriteValue
BinaryExpression	Left PrevNodeContext Right WritePos WriteType Up WriteValue PrevLeaf WriteValue
ConditionalExpression	WritePos PrevLeaf WriteValue PrevLeaf WriteValue PrevLeaf PrevNodeType WriteValue
LogicalExpression	Left WriteType Up PrevDfs WriteValue PrevLeaf WriteValue PrevLeaf WriteValue
WhileStatement	WritePos PrevDfs PrevLeaf WriteValue PrevLeaf WriteValue PrevLeaf WriteValue
DoWhileStatement	PrevDfs WriteType PrevLeaf Up PrevLeaf WriteValue PrevLeaf WriteValue WritePos
TryStatement	Left WriteType PrevLeaf PrevLeaf WriteValue
CatchClause	WritePos Left PrevNodeType WritePos WriteValue PrevLeaf WriteValue
ForStatement	WritePos PrevLeaf WriteValue PrevLeaf WriteValue PrevNodeValue PrevLeaf WriteValue
UpdateExpression	Up PrevNodeContext WritePos DownFirst WriteType PrevLeaf WriteValue PrevLeaf WriteValue
SequenceExpression	Left WriteType Up PrevDfs WriteType Left DownFirst DownLast WriteValue PrevLeaf WriteValue
ForInStatement	WritePos Up PrevLeaf Left WriteValue PrevLeaf WriteValue PrevLeaf WriteValue
SwitchStatement	WritePos PrevDfs Up PrevLeaf WriteValue PrevLeaf PrevLeaf WriteValue
SwitchCase	WritePos PrevDfs Left DownLast WriteType PrevLeaf WriteType NextLeaf WriteValue
otherwise	Up PrevNodeValue DownLast Right WriteValue Left PrevDfs DownLast WriteValue WriteType WriteValue

Table 2. Learned conditioning program for PHOG to predict terminals (node values).

Parent non-terminal symbol	TCond program
Identifier	PrevNodeType PrevNodeType WriteValue NextLeaf WriteValue NextLeaf WriteValue Up Up Right WriteType
Property	PrevLeaf PrevNodeContext NextLeaf WriteValue Left WriteValue DownLast PrevLeaf WriteValue DownFirst PrevLeaf WriteValue
LiteralString	PrevNodeType WriteValue PrevNodeType WriteValue PrevLeaf PrevNodeType WriteValue
VariableDeclarator	PrevNodeType WriteValue PrevNodeType WriteValue PrevLeaf PrevLeaf WriteValue
LiteralNumber	Left PrevDfs Right PrevNodeContext DownLast WriteValue PrevLeaf WriteValue PrevLeaf WriteValue PrevNodeType
LiteralBoolean	PrevDfs WriteType WriteValue PrevLeaf WriteValue PrevLeaf WriteValue
UnaryExpression	Left WriteType PrevDfs WriteType WriteValue PrevLeaf WriteValue
BinaryExpression	WritePos Left PrevDfs PrevNodeContext WriteType DownFirst WriteValue PrevLeaf WriteValue
LogicalExpression	PrevDfs WriteType WriteValue PrevLeaf WriteValue PrevLeaf WriteValue
LiteralRegExp	PrevNodeType WriteValue Up PrevLeaf PrevNodeType WriteValue PrevLeaf PrevLeaf WriteValue
UpdateExpression	Left PrevNodeContext Right DownLast PrevDfs WriteType WriteValue PrevLeaf WriteValue PrevLeaf WriteValue
BreakStatement	PrevLeaf Left WriteValue
otherwise	PrevNodeValue WriteType Up PrevDfs WriteValue