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).

<table>
<thead>
<tr>
<th>PARENT NON-TERMINAL SYMBOL</th>
<th>TCond PROGRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>root</td>
<td>Empty</td>
</tr>
<tr>
<td>Program</td>
<td>Left WriteType Left DownLast WriteType NextLeaf WriteValue</td>
</tr>
<tr>
<td>ExpressionStatement</td>
<td>Up WritePos PrevNodeContext DownLast WriteType PrevNodeType NextLeaf WriteValue</td>
</tr>
<tr>
<td>CallExpression</td>
<td>PrevDfs WriteValue PrevDfs WriteValue</td>
</tr>
<tr>
<td>MemberExpression</td>
<td>WritePos Up PrevNodeContext DownFirst WriteType</td>
</tr>
<tr>
<td>Property</td>
<td>Up PrevNodeContext DownLast WriteType Left DownFirst WriteType Right PrevDfs Right WriteValue</td>
</tr>
<tr>
<td>ObjectExpression</td>
<td>WritePos Left WriteValue PrevLeaf WriteValue</td>
</tr>
<tr>
<td>VariableDeclaration</td>
<td>PrevDfs WriteType PrevDfs WriteType PrevLeafWriteValue PrevNodeType PrevLeaf WriteValue</td>
</tr>
<tr>
<td>VariableDeclarator</td>
<td>Up WriteValue PrevLeaf WriteValue PrevLeaf WriteValue WriteValue</td>
</tr>
<tr>
<td>FunctionDeclaration</td>
<td>WritePos PrevDfs WriteValue PrevNodeType WriteValue</td>
</tr>
<tr>
<td>BlockStatement</td>
<td>Left WriteType Left PrevDfs Up WriteType</td>
</tr>
<tr>
<td>AssignmentExpression</td>
<td>Left WriteType PrevNodeContext Right WriteType PrevLeaf WriteValue PrevLeaf WriteValue</td>
</tr>
<tr>
<td>ArrayExpression</td>
<td>Left WriteType PrevNodeContext Right WriteType Up DownLast PrevNodeContext DownFirst WriteType PrevLeaf WriteValue</td>
</tr>
<tr>
<td>FunctionExpression</td>
<td>PrevDfs WriteValue PrevDfs WriteValue PrevLeaf WriteValue PrevLeaf WriteValue</td>
</tr>
<tr>
<td>ReturnStatement</td>
<td>PrevLeaf Right WriteValue PrevDfs WriteValue PrevLeaf WriteValue PrevLeaf WriteValue</td>
</tr>
<tr>
<td>ArrayAccess</td>
<td>PrevDfs WriteValue PrevLeaf Type PrevDfs Right PrevNodeType DownFirst WriteValue PrevLeaf WriteValue</td>
</tr>
<tr>
<td>IfStatement</td>
<td>Left WriteType Right PrevLeaf WriteValue PrevLeaf PrevNodeType WriteValue PrevLeaf WriteValue</td>
</tr>
<tr>
<td>UnaryExpression</td>
<td>Up WriteValue PrevDfs WriteValue PrevLeaf WriteValue PrevLeaf WriteValue</td>
</tr>
<tr>
<td>ThrowStatement</td>
<td>Up PrevNodeContext DownFirst WriteType PrevLeaf WriteValue PrevLeaf PrevNodeType DownFirst WriteType PrevLeaf WriteValue</td>
</tr>
<tr>
<td>NewExpression</td>
<td>PrevDfs PrevNodeContext WriteType DownFirst DownLast Right WriteType PrevLeaf WriteValue PrevLeaf WriteValue</td>
</tr>
<tr>
<td>BinaryExpression</td>
<td>PrevDfs PrevNodeContext Right WritePos WriteValue PrevLeaf WriteValue PrevLeaf WriteValue</td>
</tr>
<tr>
<td>ConditionalExpression</td>
<td>WritePos PrevLeaf WriteValue PrevLeaf WriteValue PrevLeaf WriteValue PrevNodeType WriteValue</td>
</tr>
<tr>
<td>LogicalExpression</td>
<td>Left WriteType Up PrevDfs WriteValue PrevLeaf WriteValue PrevLeaf WriteValue</td>
</tr>
<tr>
<td>WhileStatement</td>
<td>WritePos PrevLeaf WriteValue PrevLeaf WriteValue PrevLeaf WriteValue PrevLeaf WriteValue</td>
</tr>
<tr>
<td>DoWhileStatement</td>
<td>PrevDfs WriteType PrevLeaf Up PrevLeaf WriteValue PrevLeaf WriteValue PrevLeaf WriteValue</td>
</tr>
<tr>
<td>TryStatement</td>
<td>Left WriteType PrevLeaf WriteValue PrevLeaf WriteValue</td>
</tr>
<tr>
<td>CatchClause</td>
<td>WritePos Left PrevNodeType WritePos WriteValue PrevLeaf WriteValue</td>
</tr>
<tr>
<td>ForStatement</td>
<td>WritePos PrevLeaf WriteValue PrevLeaf WriteValue PrevLeafWriteValue PrevLeaf WriteValue</td>
</tr>
<tr>
<td>UpdateExpression</td>
<td>Up PrevNodeContext WritePos DownFirst WriteType PrevLeaf WriteValue PrevLeaf WriteValue</td>
</tr>
<tr>
<td>SequenceExpression</td>
<td>Left WriteType Up PrevDfs WriteType Left DownFirst DownLast WriteType PrevLeaf WriteValue</td>
</tr>
<tr>
<td>ForInStatement</td>
<td>WritePos Up PrevLeaf Left WriteValue PrevLeaf WriteValue PrevLeaf WriteValue</td>
</tr>
<tr>
<td>SwitchStatement</td>
<td>WritePos PrevDfs Up PrevLeaf WriteValue PrevLeaf WriteValue</td>
</tr>
<tr>
<td>SwitchCase</td>
<td>WritePos PrevDfs Left DownLast WriteType PrevLeaf WriteType NextLeaf WriteValue</td>
</tr>
<tr>
<td>otherwise</td>
<td>Up PrevNodeType DownLast Right WriteValue Left PrevDfs DownLast WriteValue WriteType WriteValue</td>
</tr>
</tbody>
</table>
Table 2. Learned conditioning program for PHOG to predict terminals (node values).

<table>
<thead>
<tr>
<th>Parent Non-Terminal Symbol</th>
<th>TCCond Program</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifier</td>
<td>PrevNodeType PrevNodeType WriteValue NextLeaf WriteValue NextLeaf WriteValue Up Up Right WriteType</td>
</tr>
<tr>
<td>Property</td>
<td>PrevLeaf PrevNodeType PrevNodeType WriteValue PrevNodeType PrevNodeType WriteValue Left WriteValue DownLast PrevLeaf WriteValue WriteValue DownFirst PrevLeaf WriteValue</td>
</tr>
<tr>
<td>LiteralString</td>
<td>PrevNodeType WriteValue PrevNodeType PrevNodeType WriteValue PrevLeaf PrevNodeType WriteValue</td>
</tr>
<tr>
<td>VariableDeclarator</td>
<td>PrevNodeType WriteValue PrevNodeType WriteValue PrevLeaf PrevLeaf WriteValue</td>
</tr>
<tr>
<td>LiteralNumber</td>
<td>Left PrevDfs Right PrevNodeType PrevNodeType WriteValue PrevLeaf WriteValue PrevLeaf WriteValue PrevLeaf WriteValue PrevNodeType</td>
</tr>
<tr>
<td>LiteralBoolean</td>
<td>PrevDfs WriteType WriteValue PrevLeaf WriteValue PrevLeaf WriteValue</td>
</tr>
<tr>
<td>UnaryExpression</td>
<td>Left WriteType PrevDfs WriteType WriteValue PrevLeaf WriteValue</td>
</tr>
<tr>
<td>BinaryExpression</td>
<td>WritePos Left PrevDfs PrevNodeType WriteType WriteType DownFirst WriteValue PrevLeaf WriteValue WriteValue</td>
</tr>
<tr>
<td>LogicalExpression</td>
<td>PrevDfs WriteType WriteValue PrevLeaf WriteValue PrevLeaf WriteValue</td>
</tr>
<tr>
<td>LiteralRegExp</td>
<td>PrevNodeType WriteValue Up PrevLeaf PrevNodeType WriteValue PrevLeaf WriteValue</td>
</tr>
<tr>
<td>UpdateExpression</td>
<td>Left PrevNodeType PrevNodeType Right DownLast PrevDfs WriteType WriteValue PrevLeaf WriteValue</td>
</tr>
<tr>
<td>BreakStatement</td>
<td>PrevLeaf Left WriteValue</td>
</tr>
<tr>
<td>otherwise</td>
<td>PrevNodeType WriteType Up PrevDfs WriteValue</td>
</tr>
</tbody>
</table>