Supplementary Material

Table 4. Edge types used in Augmented ASTs. The initial AST is constructed using the `AST` and `NEXT_TOKEN` edges, and then the remaining edges are added. In other words, the the "AST" model from Table 1 uses a graph that contains only the `AST` and `NEXT_TOKEN` edge types (and `WORD_USE` if it also uses a GSC), while the "AugAST" model contains all the edge types below. The reversed version of every edge is also added as its own type (e.g. `reverse_AST`, `reverse_LAST_READ`) to let the GNN message passing occur in both directions.

Edge Name	Description
`AST`	The edges used to construct the original AST.
`NEXT_TOKEN`	Edges added to the original AST that specify the left–to–right ordering of the children of a node in the AST. These edges are necessary since ASTs have ordered children, but we are representing the AST as a directed multigraph.
`COMPUTED_FROM`	Connects a node representing a variable on the left of an equality to those on the right. (E.g. edges from `y` to `x` and `z` to `x` in `x = y + z`.) The same as in (Allamanis et al., 2018).
`LAST_READ`	Connects a node representing a usage of a variable to all nodes in the AST at which that variable's value could have been last read from memory. The same as in (Allamanis et al., 2018).
`LAST_WRITE`	Connects a node representing a usage of a variable to all nodes in the AST at which that variable's value could have been last written to memory. The same as in (Allamanis et al., 2018).
`RETURNS_TO`	Points a node in a return statement to the node containing the return type of the method. (E.g. `x` in `return x` gets an edge pointing to `int` in `public static int getX(x)`.)
`LAST_SCOPE_USE`	Connects a node representing a variable to the node representing the last time this variable's name was used in the text of the code (i.e. capturing information about the text, not the control flow), but only within lexical scope. This edge exists to try and give the non–GSC models as much lexical information as possible to make them as comparable with the GSC model.
`LAST_FIELD_LEX`	Connects a field access (e.g. `this.whatever` or `Foo.whatever`) node to the last use of this.whatever (or to the variable's initialization, if it's the first use). This is not lexical–scope aware (and, in fact, can't be in Java, in general).
`FIELD`	Points each node representing a field access (e.g. `this.whatever`) to the node where that field was declared.
`WORD_USE`	Points cache nodes to nodes representing variables in which the vocab word was used in the variable's name.

Table 5. Repositories used in experiments. All were taken from the Maven repository (https://mvnrepository.com/). Entries are in the form "group/repository name/version".

Seen Repos
com.fasterxml.jackson.core/jackson-core/2.9.5
com.h2database/h2/1.4.195
javax.enterprise/cdi-api/2.0
junit/junit/4.12
mysql/mysql-connector-java/6.0.6
org.apache.commons/commons-collections4/4.1
org.apache.commons/commons-math3/3.6.1
org.apache.commons/commons-pool2/2.5.0
org.apache.maven/maven-project/2.2.1
org.codehaus.plexus/plexus-utils/3.1.0
org.eclipse.jetty/jetty-server/9.4.9.v20180320
org.reflections/reflections/0.9.11
org.scalacheck/scalacheck_2.12/1.13.5
org.slf4j/slf4j-api/1.7.25
org.slf4j/slf4j-log4j12/1.7.25
Unseen Repos
org.javassist/javassist/3.22.0-GA
joda-time/joda-time/2.9.9
org.mockito/mockito-core/2.17.0

Table 6. Accuracy (and top–5 accuracy) on the Fill–In–The–Blank task, depending on which type of GNN the model uses. See Table 1 for explanations of the abbreviations. All models use AugAST as their code representation.

		GGNN	DTNN	RGCN
Seen repos	Closed Vocab	0.80 (0.90)	0.72 (0.84)	0.80 (0.90)
	GSC	**0.97** (0.99)	0.89 (0.95)	0.95 (0.98)
Unseen repos	Closed Vocab	0.59 (0.78)	0.46 (0.68)	0.62 (0.79)
	GSC	**0.92** (0.96)	0.80 (0.89)	0.88 (0.95)

Table 7. Accuracy (and top–5 accuracy) on the Variable Naming task, depending on which type of GNN the model uses. See Table 1 for explanations of the abbreviations. All models use AugAST as their code representation.

		GGNN	DTNN	RGCN
Seen repos	Closed Vocab	0.19 (0.26)	0.23 (0.31)	0.27 (0.34)
	GSC	**0.53** (0.69)	0.33 (0.48)	0.46 (0.63)
Unseen repos	Closed Vocab	0.04 (0.07)	0.06 (0.08)	0.06 (0.09)
	GSC	**0.41** (0.57)	0.25 (0.40)	0.35 (0.49)

Table 8. Extra information about performance on the Variable Naming task. Entries in this table are of the form "subword accuracy, edit distance, edit distance divided by real name length". The edit distance is the mean of the character–wise Levenshtein distance between the produced name and the real name.

		Closed Vocab	CharCNN	Pointer Sentinel	GSC (ours)
Seen repos	AST	0.30, 7.22, 0.94	0.28, 8.67, 1.08	0.32, 8.00, 1.07	0.56, 3.87, 0.39
	AugAST	0.26, 7.64, 0.94	0.28, 7.46, 0.99	0.35, 6.51, 0.77	0.60, 3.68, 0.37
Unseen repos	AST	0.09, 8.66, 1.23	0.10, 8.82, 1.12	0.13, 9,39, 1.37	0.42, 4.81, 0.59
	AugAST	0.09, 8.34, 1.14	0.10, 8.16, 1.12	0.14, 8.03, 1.07	0.48, 4.28, 0.49