

Large Language Models and Simple, Stupid Bugs

Kevin Jesse

UC Davis

Davis, USA

krjesse@ucdavis.edu

Toufique Ahmed

UC Davis

Davis, USA

tfahmed@ucdavis.edu

Premkumar T. Devanbu

UC Davis

Davis, USA

ptdevanbu@ucdavis.edu

Emily Morgan

UC Davis

Davis, USA

eimorgan@ucdavis.edu

発表者: 中谷祐太

SStuB(Simple Stupid Bugs)

単一のステートメントかつテンプレートに一致するバグ

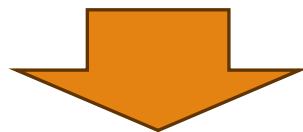
Bug → `paintBorder.top >= 1`
`paintBorder.top > 1` ← **Fix**

SStuBの例

- 16種のテンプレートに分類
 - CHANGE_OPERAND
 - CHANGE_IDENTIFIER
 - SWAP_ARGUMENTS
 - 他
- バグの修正は容易だが見つけづらい

Codex

- GitHubの公開リポジトリなど,利用可能なソースからの自然言語とソースコードで構成
- 主にPythonで学習
- コード補完,プログラムの修復, 数学など様々なタスクで機能する
- 広範なベンチマーク,セキュリティの脆弱性チェック,経験的評価を受けている



特殊なバグであるSStuBについては評価されていない

目的

- Codexと他のコード補完モデルがSStuBを生成するかどうか,または正しく修正するかを評価する

コード補完モデル

PolyCoder

- GPT-2アーキテクチャに基づいている
- C,C++,Javaを主要言語とする12のプログラミング言語で学習

CodeGen

- GitHubのコードと自然言語のコーパスで学習
- 複数の言語で学習しているが主にPythonで学習

ManySStuBs4Jデータセット

- MavenプロジェクトからマイニングされたSStuBから構成されたJavaのデータセット
- 63,923個のSStuBからなる

フィルタリング

1. LLMにバイアスを与えるようなSStuBを除去
2. 重複しているSStuBを除去



16,899個のSStuB

実験方法

1. SStuBより前のコードをプロンプトとし、コード補完により続きを生成

プロンプト

```
JBTabImpl.java
1524     if (paintBorder.top > 0) {
1525         if (isHideTabs()) {
1526             if (isToDrawBorderIfTabsHidden()) {
1527                 g2d.setColor(borderColor);
1528                 g2d.fill(shaper.reset().doRect(boundsX, boundsY, boundsWidth,
1529                     1).getShape());
1530             }
1531         }
1532         else if (isStealthModeEffective()) {
1533             //g2d.setColor(borderColor);
1534             //g2d.fill(shaper.reset().doRect(boundsX, boundsY - 1,
1535                 // boundsWidth, 1).getShape());
1536         }
1537         else {
1538             Color tabFillColor = getActiveTabFillIn();
1539             if (tabFillColor == null) {
1540                 tabFillColor = shape.path.transformY1(shape.to, shape.from);
1541             }
1542             g2d.setColor(tabFillColor);
1543             g2d.fill(shaper.reset().doRect(boundsX, topY + shape.path.deltaY(1),
1544                 boundsWidth, paintBorder.top).getShape());
1545             //If the top of the border is a non-paint border, then the border is painted.
1546             if (
1549                 paintBorder.top >= 1
1550                 paintBorder.top > 1 ← Fix

```

2. 生成された文とSStuB及び修正後のコードを比較
3. 生成された文をバグ,パッチ,一致せずの3種に分類

RQ1:LLMのSStuB生成頻度

結果

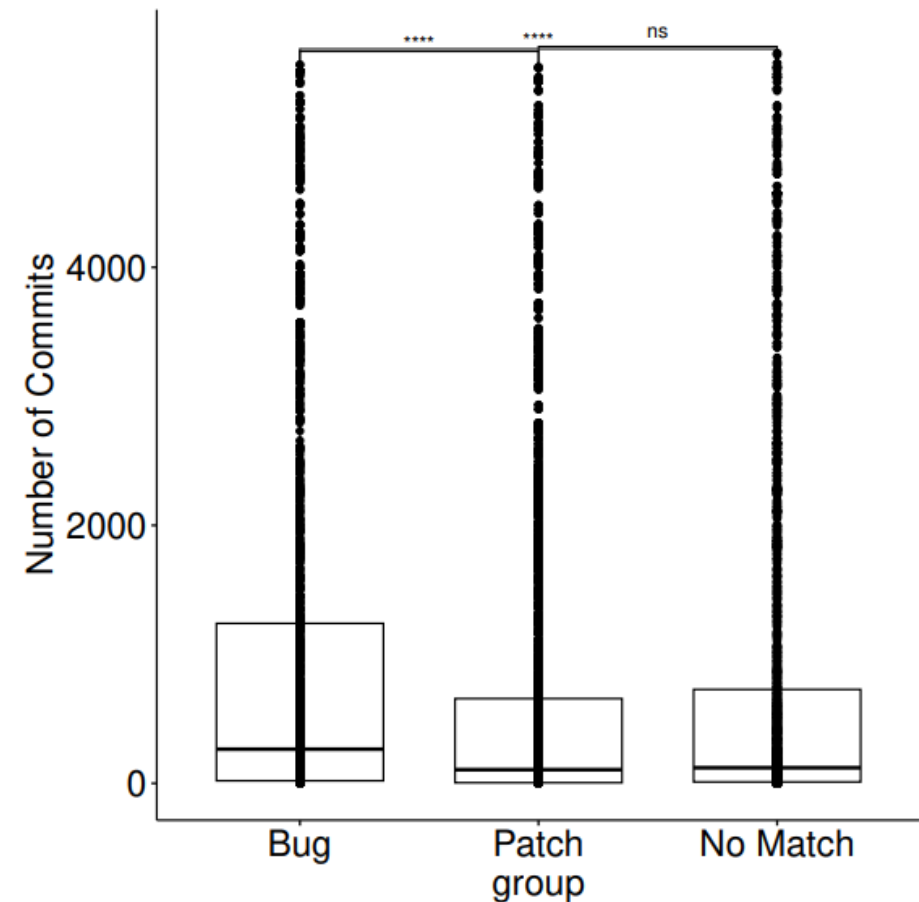
正しいコードに対し2倍近くのものSStuBを生成

Codexは13.41%のSStuBを回避

Model	Bugs	Patches	No Match	Bug/Patch	Match Rate (%)
PolyCoder 160M	3429	1635	11835	2.10	29.97
PolyCoder 0.4B	3672	1852	11375	1.98	32.69
PolyCoder 2.6B	3924	2096	10879	1.87	35.62
CodeGen 350	3709	1911	11279	1.94	33.26
CodeGen 2B	4102	2756	10041	1.49	40.58
CodeGen 6B	4168	2944	9787	1.42	42.09
CodeGen 16B	4299	3296	9304	1.30	44.94
Cushman 12B	3775	1833	11291	2.06	33.19
Davinci 175B	4452	2267	10180	1.96	39.76

RQ2: Codexが生成したSStuBの修正時間

CodexがSStuBを生成した場合,それ以外の場合よりも修正に時間がかかる可能性がある

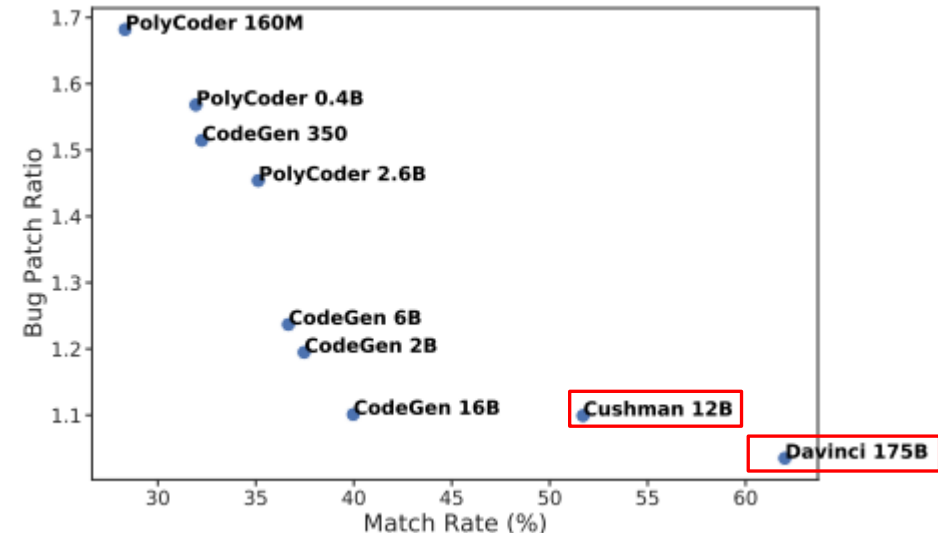
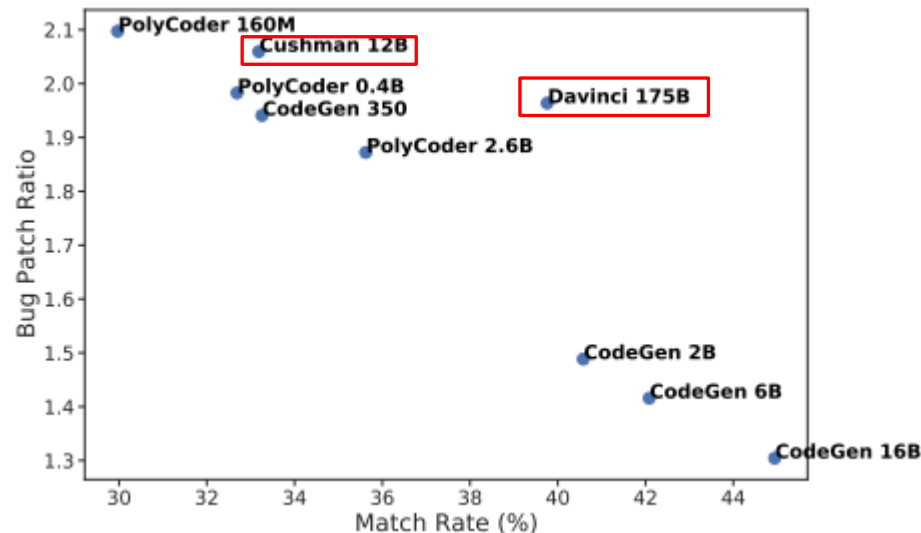


RQ4:プロンプトに自動生成コメントを追加した場合のSStuB生成頻度

パッチからCodeTransモデルでコメントを生成し,再び実験を行った。

結果

- コメントによりバグ/パッチ比率が改善する
- Codexはコメントにより著しく改善する



RQ5:プロンプトにバグ由来のコメントを追加した場合のSStuB生成頻度

SStuBからCodeTransモデルでコメントを生成し,再び実験を行った

結果

- Codexはバグから生成されたコメントでもバグ/パッチ比率が改善する

Model Name	Model Size (Billions)	Bug Change	% Change	Patch Change	% Change	Bug/Patch Ratio Change	% Change	Match Rate Change
PolyCoder 160M	0.16	59	1.72	-237	-14.50	0.40	18.97	-1.05
PolyCoder 400M	0.4	171	4.66	-132	-7.13	0.25	12.69	0.23
PolyCoder 2.6B	2.6	234	5.96	-145	-6.92	0.26	13.84	0.53
CodeGen 350M	0.35	110	2.97	-178	-9.31	0.26	13.54	-0.40
CodeGen 2.7	2.7	-78	-1.90	-420	-15.24	0.23	15.74	-2.95
CodeGen 6B	6.1	-264	-6.33	-603	-20.48	0.25	17.79	-5.13
CodeGen 16B	16.1	-95	-2.21	-558	-16.93	0.23	17.72	-3.86
Codex Cushman	12.0	1511	40.03	1135	61.92	-0.28	-13.52	15.66
Codex Davinci	175.0	1620	36.39	1613	71.15	-0.40	-20.31	19.13

まとめ

- 目的

Codexと他のコード補完モデルがSStuBを生成するかどうか,または正しく修正するかを評価する

- 結果

- LLMを用いることでSStuBを一部回避できる
- 同時により多くのSStuBを発生させる可能性がある
- コメントによってバグ/パッチ比率を改善できる
- Codexはコメントの厳密な正しさを必要としない

RQ3: SStuBの規則性

バグとパッチの負の対数尤度にほとんど差がない

