

提升预训练生成模型在第三方代码库的性能

CERT: Continual Pre-Training on Sketches for Library-Oriented Code Generation (IJCAI2022)

咎道广(daoguang@iscas.ac.cn), 关贝, 王永吉等

科研背景:

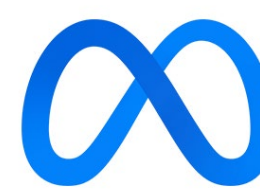
随着预训练语言模型的快速发展, 该技术也被应用到了代码生成领域。OpenAI的Codex, DeepMind的AlphaCode等模型相继被提出用于自动生成代码, 并表现出了优异的性能。然而这一类模型均没有开源, 所以我们首先研发了PyCodeGPT-110M, 它能够在OpenAI提供的标准数据集HumanEval上和同等级别参数大小的Codex性能相媲美, 然后我们基于这一强大的代码生成模型PyCodeGPT去提升了面向第三方代码库的代码生成性能。



AlphaCode



Codex



InCoder

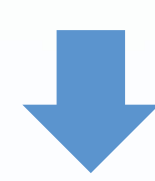


CodeGen



PyCodeGPT

```
import pandas as pd
# reading data from anyonijcai22.csv
into DataFrame w/the argument
`names`: a list ['user_1', 'user_2',
'user_3']
df =
```



```
pd.read_csv(anonijcai22.csv,
names=[user_1, user_2, user_3])
```

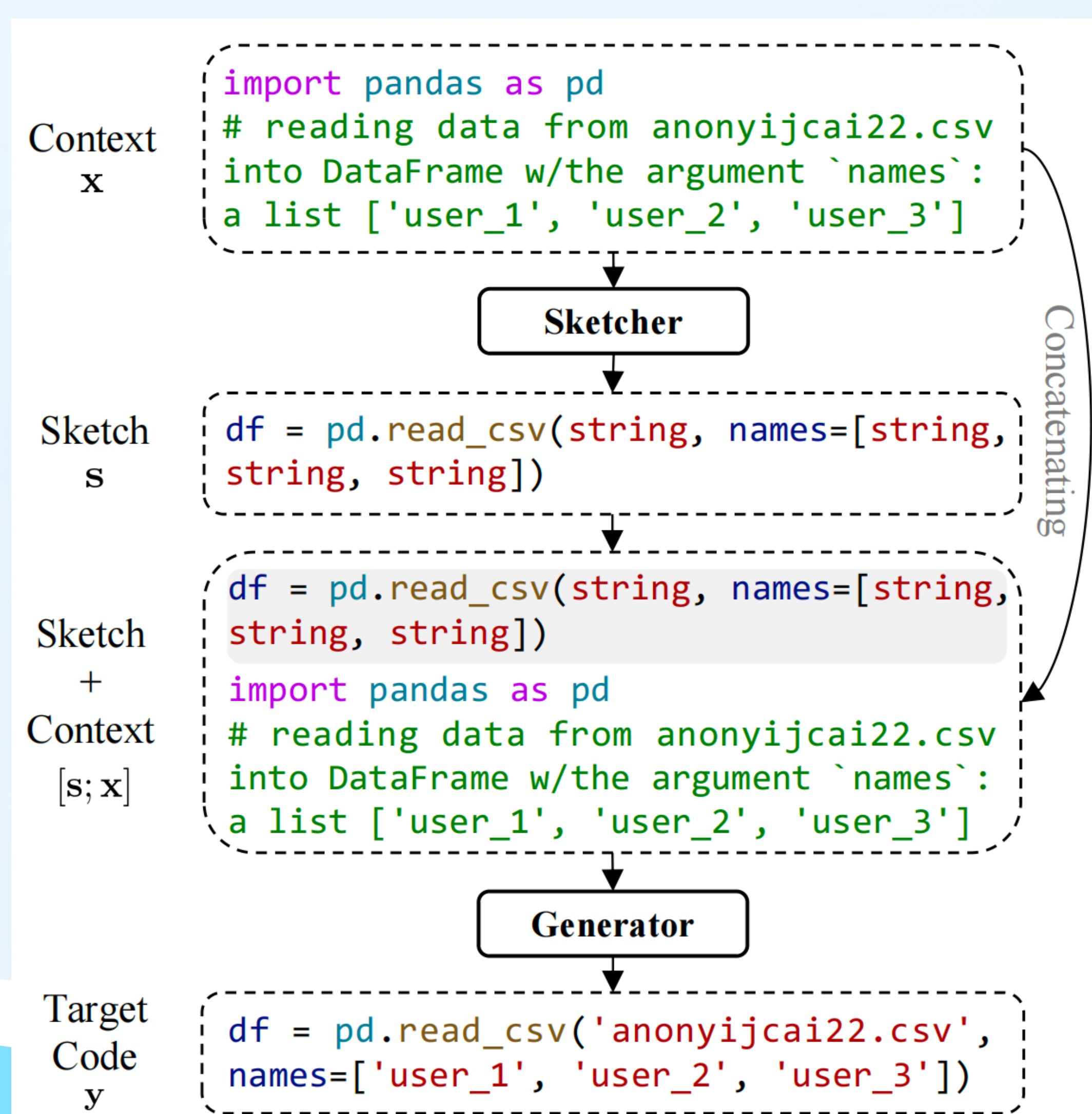
解决的科学难题:

- 目前先进的代码生成模型均没有开源, 如何训练一个性能优异的代码生成模型?
- 提升Codex、PyCodeGPT这些通用的代码生成性能在第三方代码库的代码生成性能。
- 如何评价面向第三方代码库的代码生成性能?

创新点:

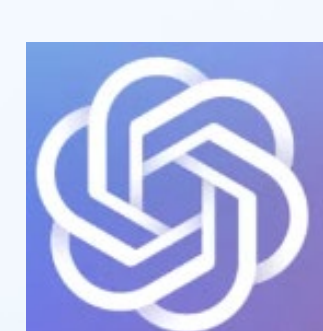
- 提出了基于Python的代码生成模型PyCodeGPT。
- 在大量无监督代码语料上利用代码骨架的思想进一步提升了面向第三方代码库的代码生成性能, 提出了CERT (SketCher和GenERaTor)。
- 构建了两个第三方代码库的标准数据集: PandasEval和NumpyEval。

解决方案:

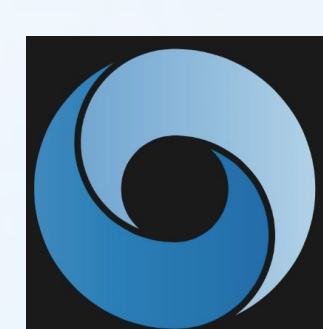


CERT的框架: Sketcher为编程问题context生成代码骨架sketch, 然后Generator为编程问题context和代码骨架sketch生成目标代码target code。

ISCAS PyCodeGPT



Codex



AlphaCode

Continual Pre-Training

CERT

实验性能:

Model	Params	pass@k		
		k=1	k=10	k=100
TabNine	-	2.58%	4.35%	7.59%
GPT-Neo	125M	0.75%	1.88%	2.97%
CodeParrot	110M	3.80%	6.57%	12.78%
PolyCoder	160M	2.13%	3.35%	4.88%
Codex	85M	8.22%	12.81%	22.40%
AlphaCode	89M	4.3%	12.2%	20.0%
PyCodeGPT	110M	8.33%	13.36%	19.13%

表1: PyCodeGPT和其他模型在HumanEval上的性能指标

Bench mark	Model	pass@1	pass@10	pass@100
Pandas Eval	CodeT5	0.00%	0.00%	0.00%
	CodeGPT	0.62%	2.65%	4.95%
	CodeClippy	0.14%	0.92%	1.92%
	CodeParrot	3.21%	13.62%	33.27%
	CodeGen	14.24%	30.71%	46.04%
	CodeGen-CERT	26.40%	46.49%	58.16%
Numpy Eval	PyCodeGPT	12.75%	37.80%	59.65%
	PyCodeGPT-CERT	28.42%	48.04%	60.96%
	CodeT5	0.00%	0.10%	0.74%
	CodeGPT	1.59%	4.17%	8.54%
	CodeClippy	0.08%	0.59%	1.24%
	CodeParrot	8.42%	21.46%	45.94%
	CodeGen	19.31%	40.89%	60.58%
	CodeGen-CERT	32.00%	49.45%	67.82%
PyCodeGPT	18.04%	38.13%	63.37%	
PyCodeGPT-CERT	31.47%	46.42%	66.41%	

表2: CERT和其他模型在PandasEval和NumpyEval上的性能指标