

Retrieval-Based Neural Code Generation

Shirley Anugrah Hayati Raphaël Olivier Pravalika Avvaru
Pengcheng Yin Anthony Tomasic Graham Neubig

Carnegie Mellon University



Carnegie Mellon University
Language Technologies Institute

Code Generation

Given natural language description, generate a code.

params is an empty list

```
params = [ ]
```



Code Generation

Given natural language description, generate a code.

params is an empty list

`params = []`



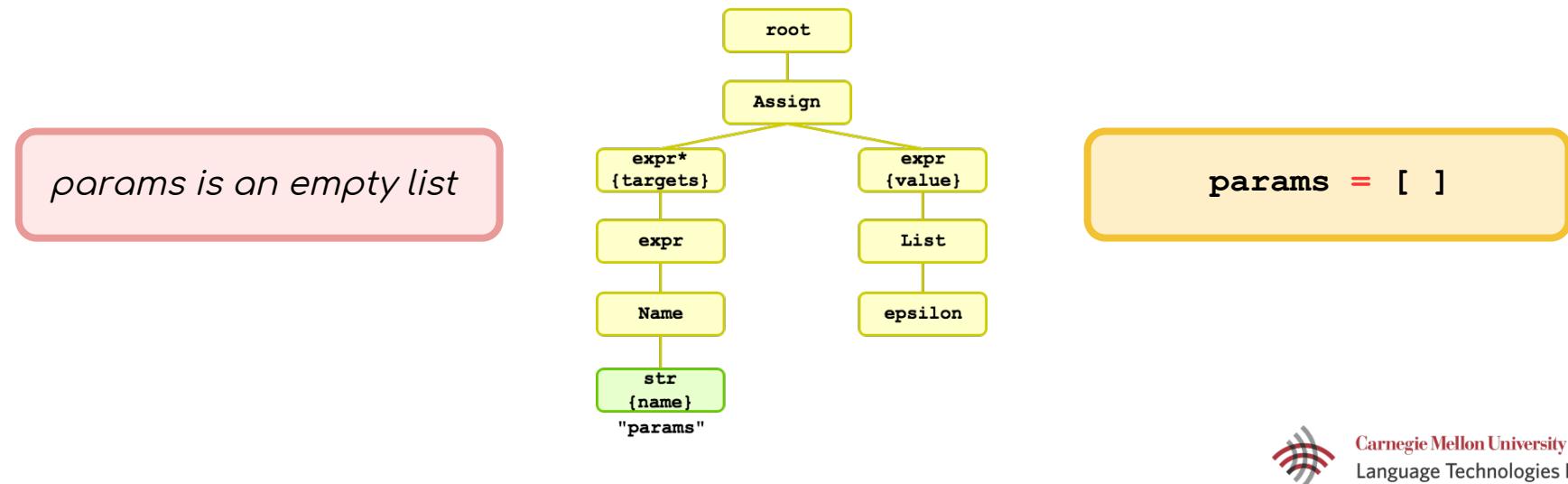
```
class DoctorBoom(MinionCard) :  
    def __init__(self):  
        super().__init__("Dr. Boom",  
        7, CHARACTER_CLASS.ALL,  
        CARD_RARITY.LEGENDARY,  
        battlecry=Battlecry(Summon(  
            BoomBot(), 2),  
            PlayerSelector()))  
  
    def create_minion(self, player):  
        return Minion(7, 7)
```



Carnegie Mellon University
Language Technologies Institute

Syntactic Code Generation (Yin+ 2017; Rabinovich+ 2017)

- Code has an underlying structure (Abstract Syntax Tree)
- Well-formed output code & improved accuracy

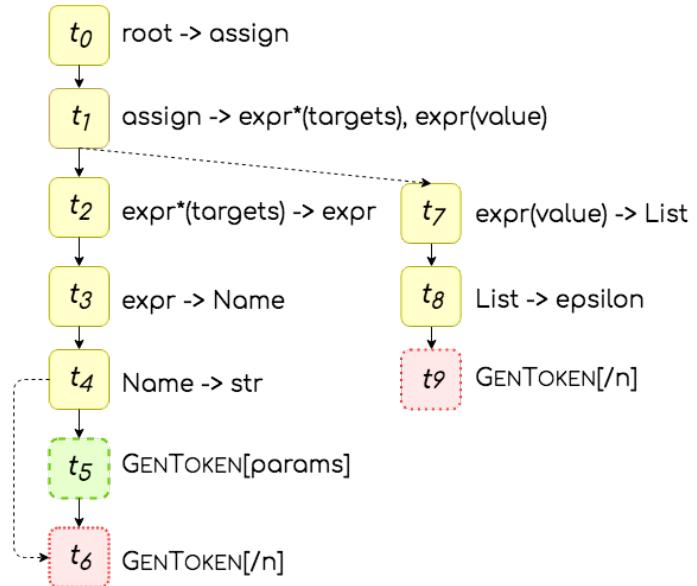


Baseline (Yin+ 2017)

Input: params is an empty list

Output: params = []

Action Tree



Baseline (Yin+ 2017)

Input: params is an empty list

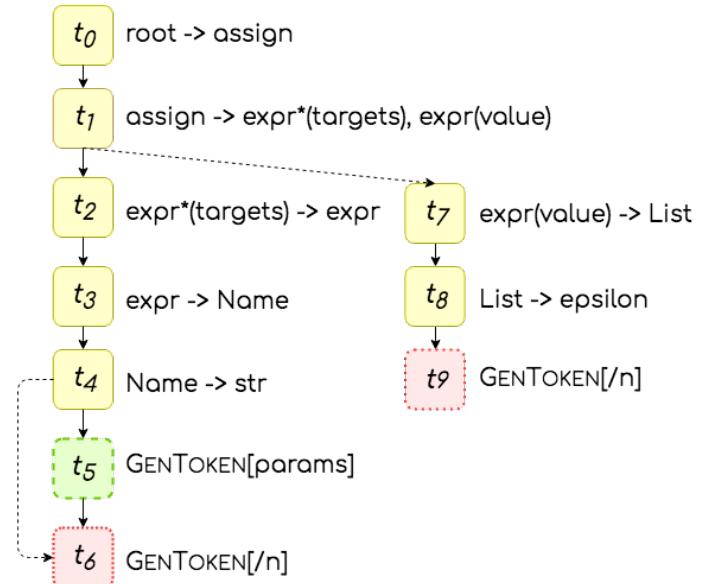
Output: params = []

Actions:

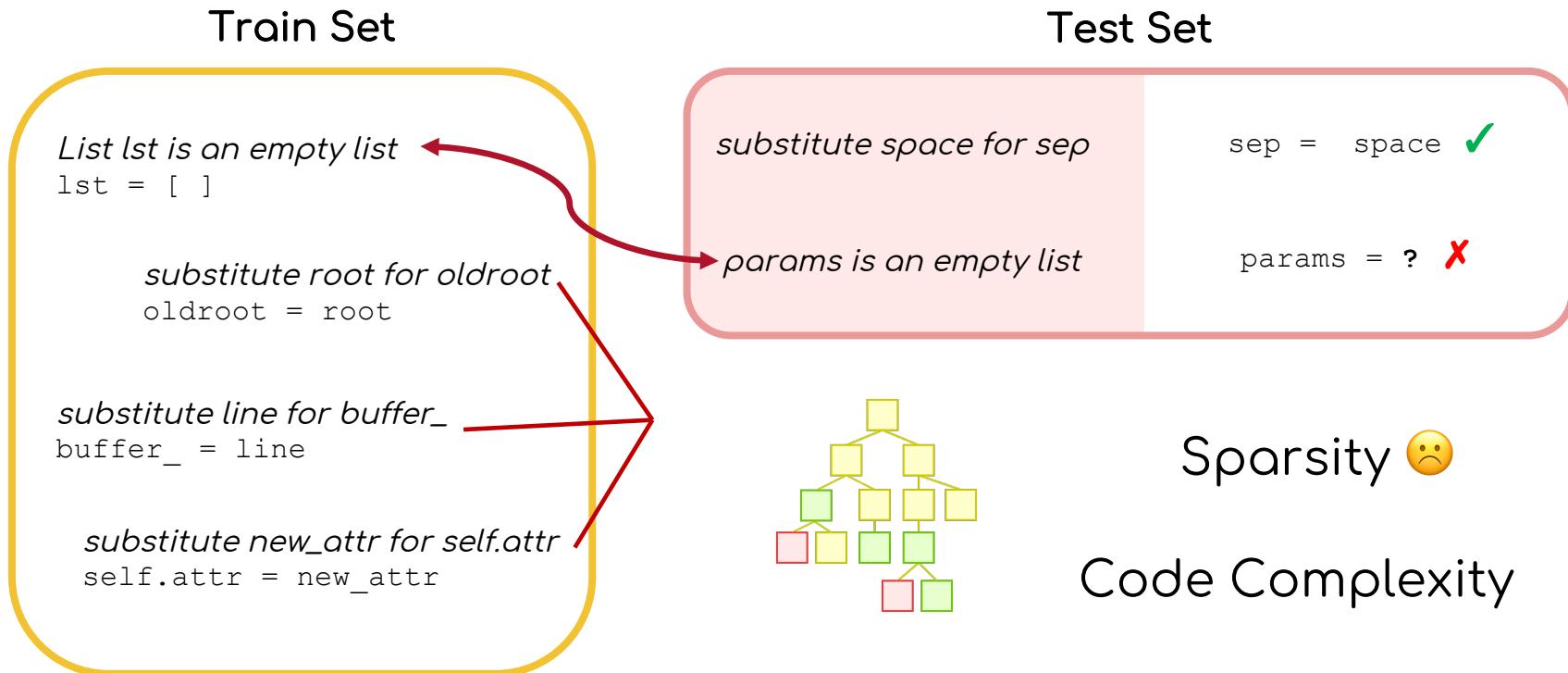
- █ Apply Rule
- █ Generate Token
- █ Generate Token with Copy

Neural Model: bidirectional Encoder-Decoder with Action Embedding, Context Vector, Parent Feeding, Copying Mechanism

Action Tree



Limitation



Codes from the Internet

Intention

Initially “params” is an empty list.

List “lst” is a list of [1, 3, 2].

Substitute “lst” for “params”.

Now sort “lst” in reversed order
so the result is [2, 3, 1].

Code

```
>>> params = [ ]  
>>> lst = [1, 3, 2]  
>>> params = lst
```



Codes from the Internet

Intention

Initially “params” is an empty list.

List “lst” is a list of [1, 3, 2].

Substitute “lst” for “params”.

Now sort “lst” in reversed order
so the result is [2, 3, 1].



Code

```
>>> params = [ ]  
>>> lst = [1, 3, 2]  
>>> params = lst
```

?



Carnegie Mellon University
Language Technologies Institute

Codes from the Internet

Intention

Initially “params” is an empty list.

List “lst” is a list of [1, 3, 2].

Substitute “lst” for “params”.

Now sort “lst” in reversed order
so the result is [2, 3, 1].

Code

```
>>> params = [ ]  
>>> lst = [1, 3, 2]  
>>> params = lst
```

```
>>> x = ['a', 'c', 'b']  
>>> x.reverse()  
>>> x  
['b', 'c', 'd']
```



Carnegie Mellon University
Language Technologies Institute

Codes from the Internet

Intention

Initially “params” is an empty list.

List “lst” is a list of [1, 3, 2].

Substitute “lst” for “params”.

Now sort “lst” in reversed order
so the result is [2, 3, 1].



Code

```
>>> params = []
```

```
>>> lst = [1, 3, 2]
```

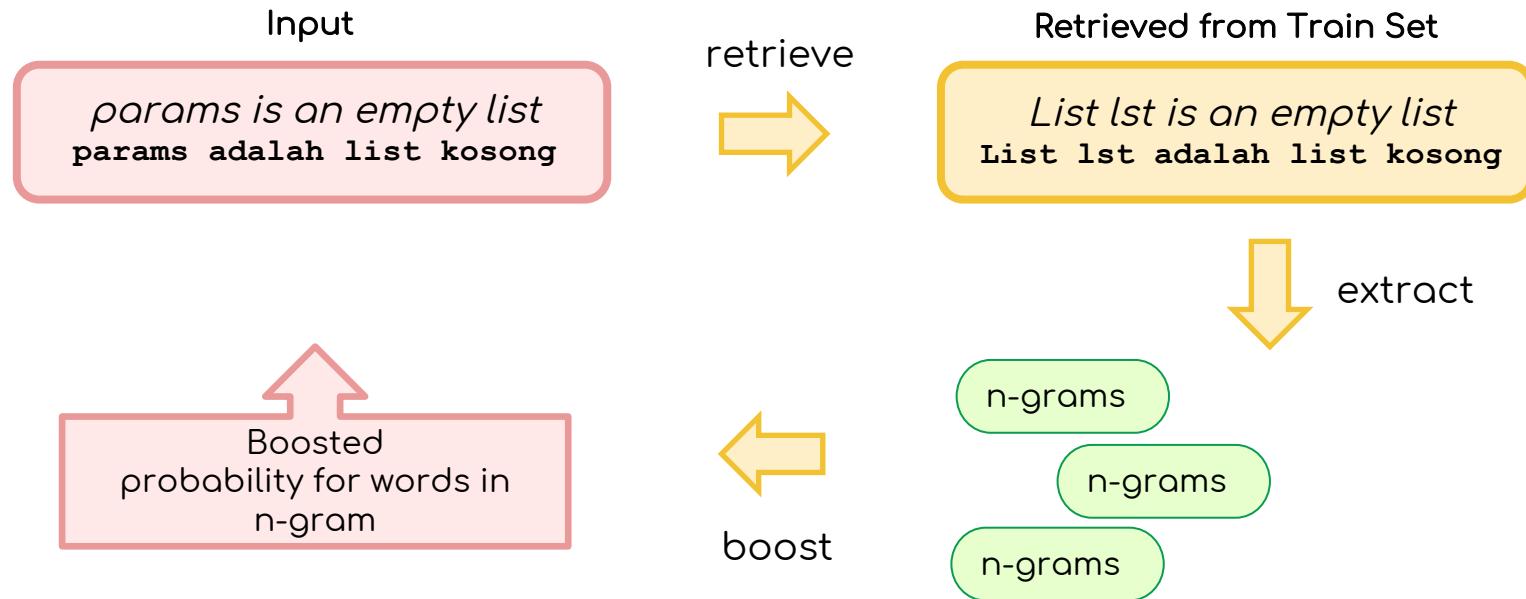
```
>>> params = lst
```

```
>>> lst.reverse()
```



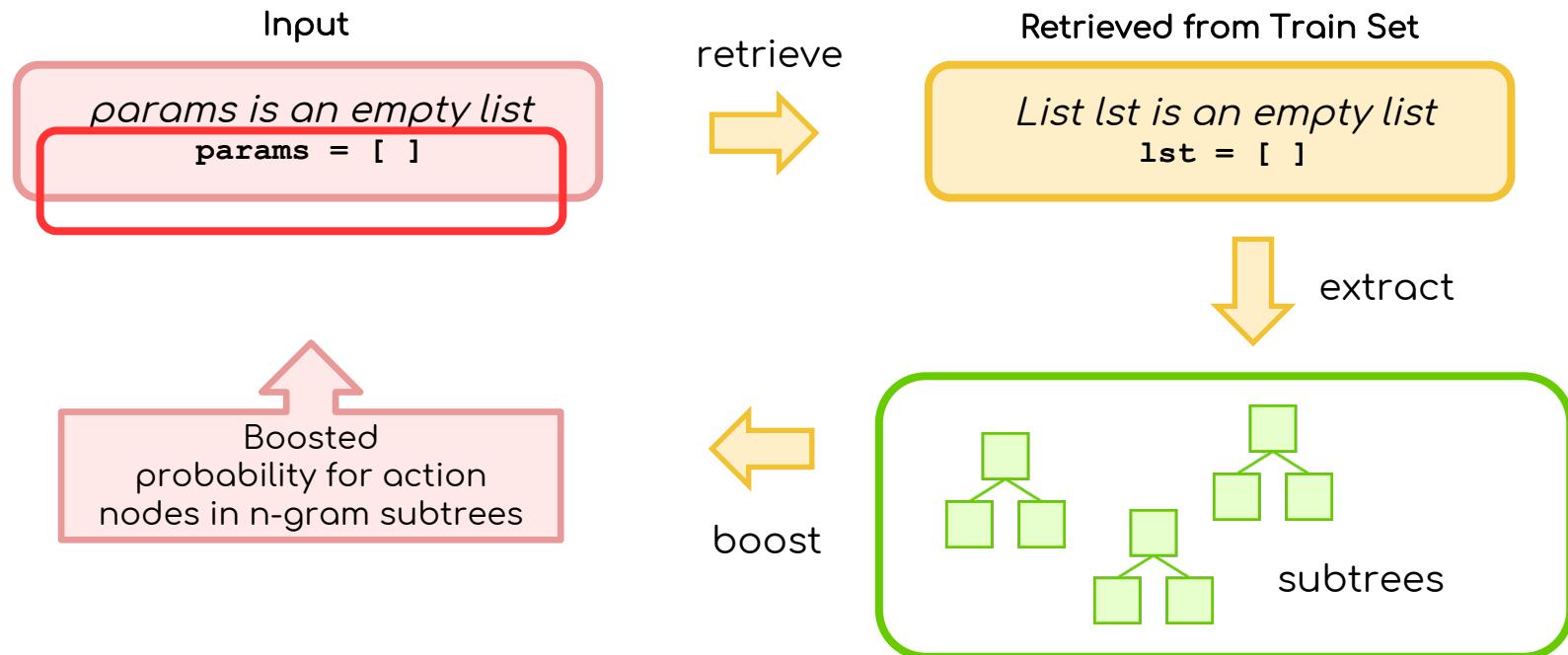
Neural Machine Translation + Retrieval

(Gu et al., 2018; Zhang et al., 2018)



Carnegie Mellon University
Language Technologies Institute

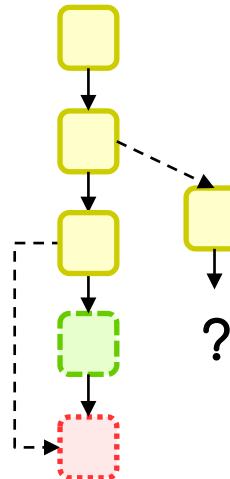
RECODE: Neural Code Generation + Retrieval



RECODE Example

Input

params is an empty list

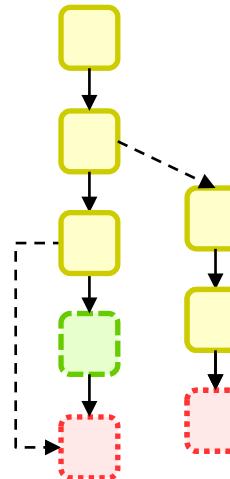


params = []



Retrieved from Train Set

List lst is an empty list

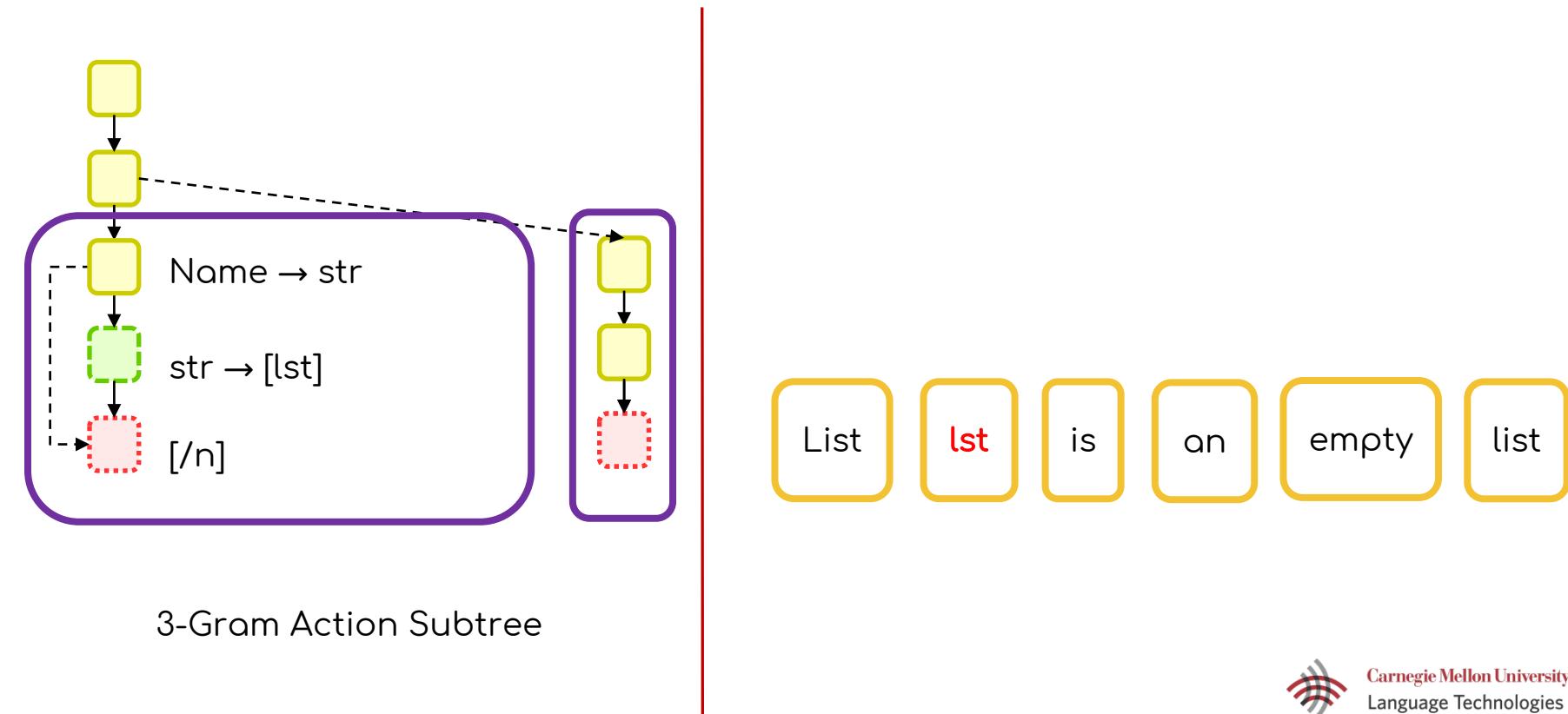


lst = []

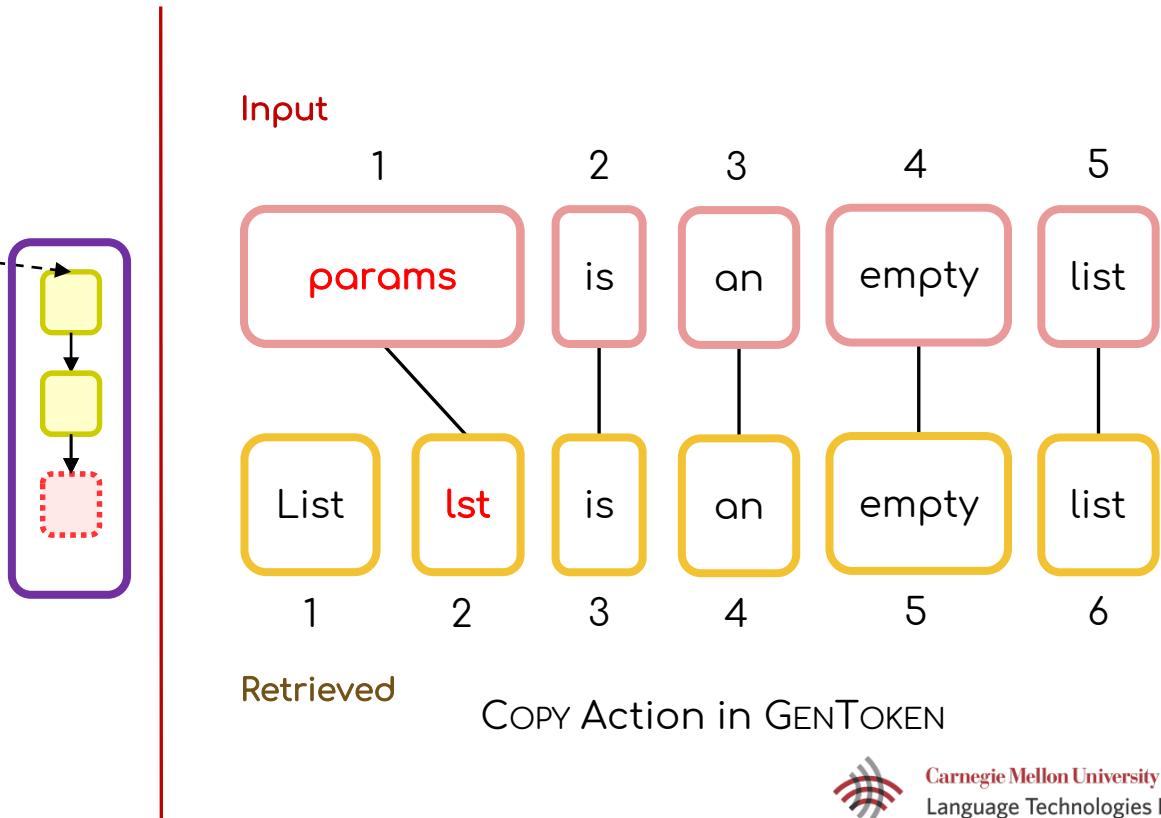
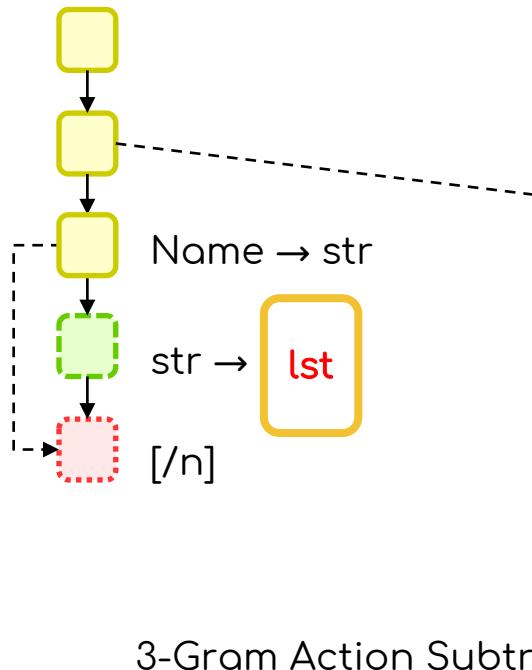


Carnegie Mellon University
Language Technologies Institute

N -gram Action Subtree



N -gram Action Subtree



Dataset

Hearthstone

(Ling et al., 2016)



*NAME_BEGIN Dr. Boom NAME_END
ATK_BEGIN 7 ATK_END*



```
class DoctorBoom(MinionCard):  
    def __init__(self):  
        super().__init__("Dr. Boom",  
                        7, CHARACTER_CLASS.ALL,  
                        CARD_RARITY.LEGENDARY, ...)
```



665 instances

Django

(Oda et al., 2015)



params is an empty list



```
params = [ ]
```



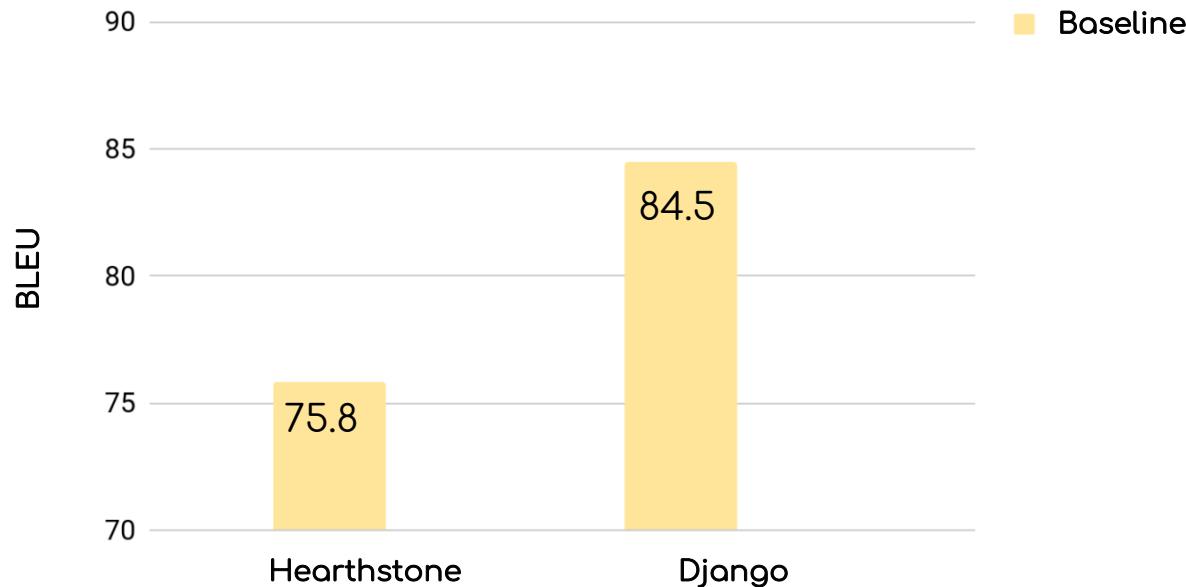
18,805 instances



Carnegie Mellon University
Language Technologies Institute

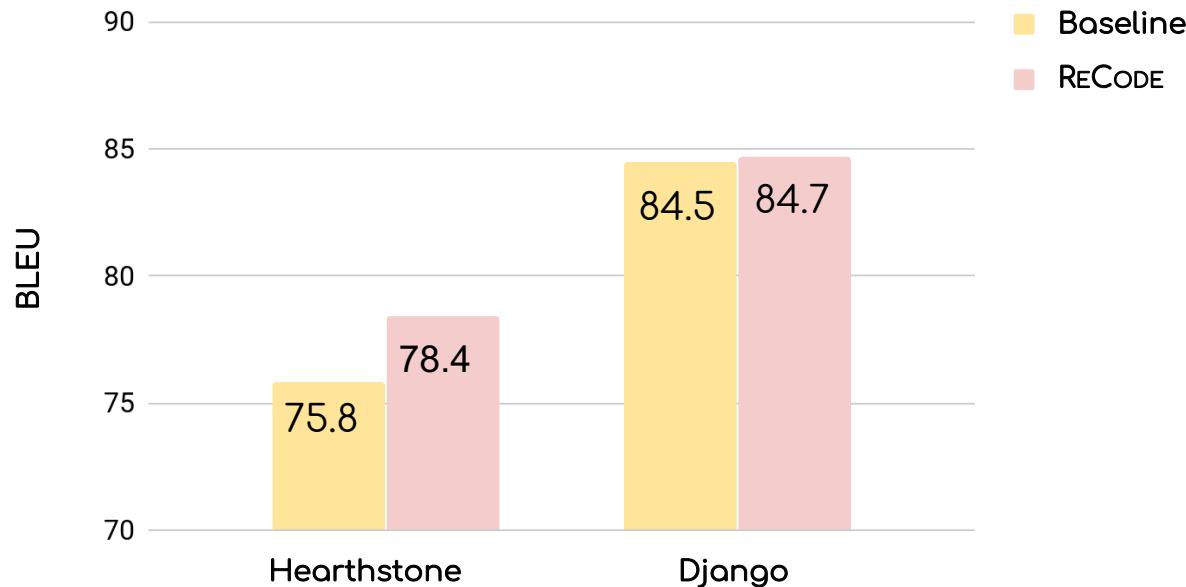
Result (BLEU Score)

All improvements are statistically significant with $\rho < 0.001$



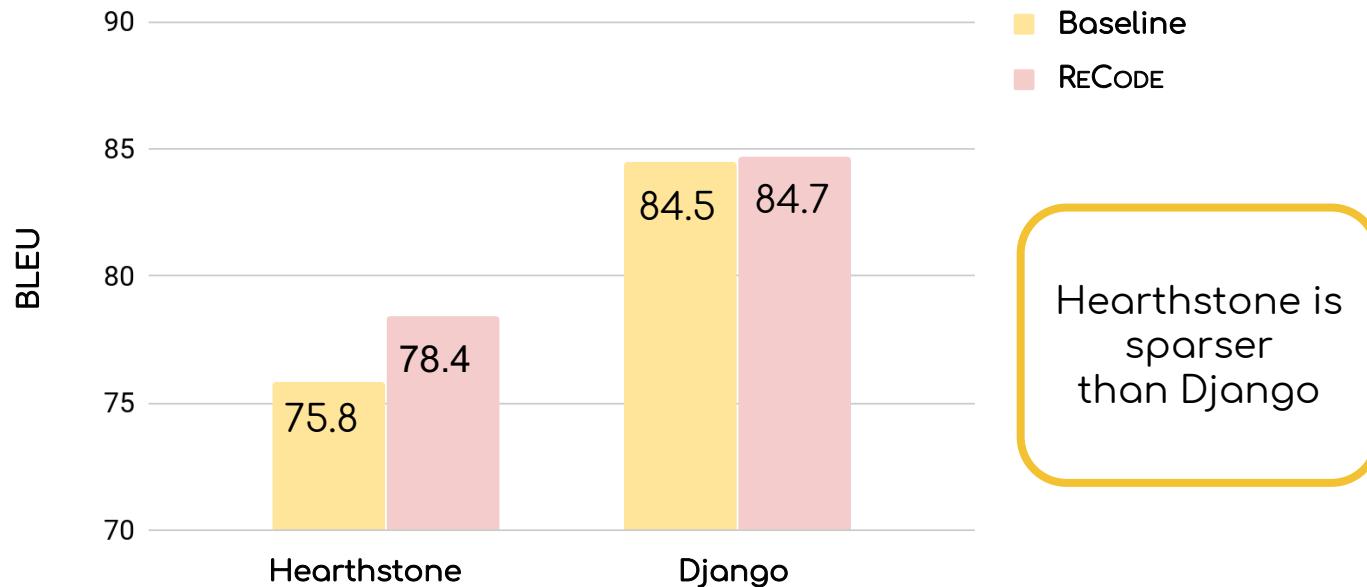
Result (BLEU Score)

All improvements are statistically significant with $\rho < 0.001$



Result (BLEU Score)

All improvements are statistically significant with $\rho < 0.001$



Example

“if offset is lesser than integer 0, sign is set to ‘-’, otherwise sign is ‘+’ ”

```
sign = ' -' if offset < 0 else '+'           Gold
```

```
sign = offset < 0 or '+'                      Baseline
```

```
sign = ' -' if offset < 0 else '+'           RECODE
```



Summary and Future Work

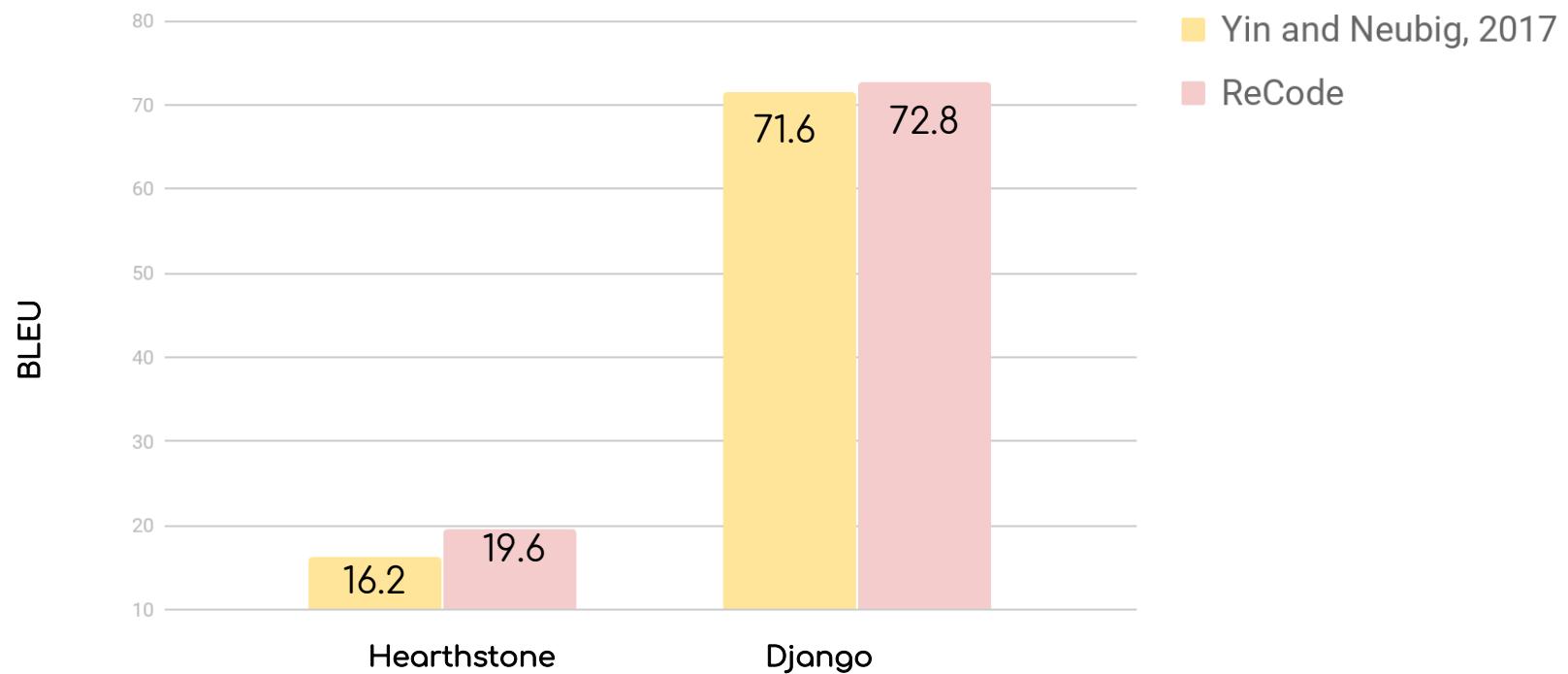
- Retrieval method for tree-based code generation model
- Effective for dataset with sparsity
- Define n -gram action subtree
- Future work on copying variable names



Thank you!



Result (Exact Match Accuracy)



Hearthstone Example

```
NAME_BEGIN Earth Elemental NAME_END ATK_BEGIN 7      Input
ATK_END DEF_BEGIN 8 DEF_END COST_BEGIN 5
COST_END DUR_BEGIN -1 DUR_END TYPE_BEGIN Minion
TYPE_END PLAYER_CLS_BEGIN Shaman PLAYER_CLS_END
RACE_BEGIN NIL RACE_END RARITY_BEGIN Epic RARITY_END
DESC_BEGIN Taunt . Overload : ( 3 ) DESC_END.
```

```
class EarthElemental (MinionCard) :                  YN17
    def __init__ (self) :
        super().__init__("Earth Elemental", 5,
                          CHARACTER_CLASS.SHAMAN, CARD_RARITY.EPIC,
                          buffs=[Buff(ManaChange(Count
                                MinionSelector(None, BothPlayer()), -1))])
    def create_minion (self, player) :
        return Minion(7, 8, taunt=True)
```

```
class EarthElemental (MinionCard) :                  RECODE
    def __init__ (self) :
        super().__init__("Earth Elemental", 5,
                          CHARACTER_CLASS.SHAMAN, CARD_RARITY.EPIC,
                          overload=3)
    def create_minion (self, player) :
        return Minion(7, 8, taunt=True)
```

```
class EarthElemental (MinionCard) :                  Gold
    def __init__ (self) :
        super().__init__("Earth Elemental", 5,
                          CHARACTER_CLASS.SHAMAN, CARD_RARITY.EPIC,
                          overload=1)
    def create_minion (self, player) :
        return Minion(7, 8, taunt=True)
```

