PLANNING WITH LARGE LANGUAGE MODELS FOR CODE GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Existing large language model-based code generation pipelines typically use beam search or sampling algorithms during the decoding process. Although the programs they generate achieve high token-matching-based scores, they often fail to compile or generate incorrect outputs. The main reason is that conventional Transformer decoding algorithms may not be the best choice for code generation. In this work, we propose a novel Transformer decoding algorithm, Planning-Guided Transformer Decoding (PG-TD), that uses a planning algorithm to do lookahead search and guide the Transformer to generate better programs. Specifically, instead of simply optimizing the likelihood of the generated sequences, the Transformer makes use of a planner that generates complete programs and tests them on public test cases. The Transformer can therefore make more informed decisions and generate tokens that will eventually lead to higher-quality programs. We also design a mechanism that shares information between the Transformer and the planner to make our algorithm computationally efficient. We empirically evaluate our framework with several large language models as backbones on public coding challenge benchmarks, showing that 1) it can generate programs that consistently achieve higher performance compared with competing baseline methods; 2) it enables controllable code generation, such as concise codes and highly-commented codes by optimizing modified objectives[1].

1 INTRODUCTION

Large language models like Transformer (Vaswani et al., 2017a) have shown successes in natural language processing, computer vision, and other domains. Thanks to Transformer's power on sequence modeling, it has been adopted for code generation (Wang et al., 2021; Ahmad et al., 2021) by treating programs as text sequences. Transformer has achieved significant improvements on the benchmarking tasks of code translation (Roziere et al., 2022), code completion (Chen et al., 2021a), and coding challenge competence (Hendrycks et al., 2021). Recently, AlphaCode (Li et al., 2022) even achieved a competitive-level performance in programming competitions with the help of large Transformer models pre-trained on a large programming corpus.

Transformer-based pipelines like AlphaCode follow the tradition of natural language processing and use sampling methods (Fan et al., 2018; Dabre & Fujita, 2020) during the generation process. Specifically, they sample a large number of complete programs using a pretrained code generation Transformer, evaluate these programs using the public test cases provided in the dataset, and output the program that passes the most number of test cases. Compared with beam search-based methods, these sampling followed by filtering algorithms (which we will refer to as *sampling + filtering*) can take advantage of test cases and indeed improve the quality of the generated programs. However, their algorithms are not sample efficient. During the Transformer generation process, they do not consider the test cases at all. Instead, they only use the test cases to evaluate the programs after all the candidate programs are generated. Different from natural languages, programs may fail completely with even a single incorrect generated token. So these algorithms may need to exhaustively sample a large number of programs to find a correct one.

The main reason behind the sample efficiency issue of these algorithms is that the Transformer beam search algorithm and the sampling algorithm (Vaswani et al., 2017b) may not be the best

[1]Project page: https://codeaimcts.github.io

choices for code generation. An ideal code generation algorithm should stop early in the generation process when it knows the program it currently generates would certainly fail and bias the generation process towards generating successful programs that pass more test cases. To achieve such a goal, we contribute to applying a planning algorithm in the Transformer generation process. Since a planning algorithm can use the pass rates of the generated programs as its objective, we use it to determine the quality of the generated codes and make the Transformer model make more informed decisions.

In this paper, we investigate the following research question: *Can we integrate a planning algorithm with a pretrained code generation Transformer, resulting in an algorithm that generates better programs than the conventional Transformer generation algorithms and the well-accepted sampling + filtering scheme in the literature?* To answer this question, we propose a novel algorithm, *Planning-Guided Transformer Decoding* (PG-TD). In this algorithm, we have a backbone Transformer that is pretrained on code generation. In each step that the Transformer generates a token, the planner does lookahead search and suggests the token that passes the most number of test cases. The planner alone may not find high-quality codes due to the large space of possible codes. So the Transformer beam search algorithm and the next-token probabilities are used inside the planner to provide useful heuristics. We find a straightforward integration between the planner and the Transformer can be computationally inefficient. So we design mechanisms that allow the Transformer and the planner to share their information to make the overall algorithm more efficient.

We emphasize that our algorithm is *model-agnostic*, that is, any standard code generation Transformer model can be used as the backbone Transformer. Importantly, our algorithm does not require acquiring more sample solutions or finetuning the Transformer model to improve its performance. We empirically find that our proposed algorithm generates higher-quality programs under multiple accepted metrics compared with competing baseline methods. Additionally, we also empirically show that our algorithm has the following advantages. 1) By changing the reward function of the planner, our algorithm becomes versatile and can optimize different objective functions without the necessity of finetuning the Transformer model. 2) Our algorithm can generate solutions that are used to finetune a code generation Transformer model to improve the Transformer's performance. More precisely, we have the following contributions in this paper.

- First, we propose a novel algorithm, Planning-Guided Transformer Decoding (PG-TD), that uses a planning algorithm for lookahead search and guide the Transformer to generate better codes. Our algorithm is model-agnostic, which can work with any standard Transformer model, and does not require knowledge of the grammar of the generated programs.
- Second, a direct integration of the planning algorithm into the Transformer decoding function can cause redundant uses of the Transformer beam search algorithm. We contribute to designing mechanisms that significantly improve the computational efficiency of the algorithm.
- Third, we evaluate our algorithm on competitive programming benchmarks and empirically show that our algorithm can consistently generate better programs in terms of the pass rate and other metrics compared with the baseline methods. We also show that our algorithm is versatile and can optimize objectives other than the pass rate for controllable code generation, such as concise codes and highly-commented codes.

2 Related Work

Transformers for program synthesis. Our work is based on Transformer for program synthesis (Roziere et al., 2020; Austin et al., 2021). Inspired by their impressive capacities on a range of natural language tasks, modern transformer-based language models (Devlin et al., 2019; Radford et al., 2019; Raffel et al., 2020) have been adopted for program synthesis by treating programming languages in the same way as natural languages. A family of BERT-based Transformers are developed to code syntax (Kanade et al., 2020; Feng et al., 2020; Devlin et al., 2019; Guo et al., 2020). Later, CodeX (Chen et al., 2021a) and CodeT5 (Wang et al., 2021) adopted GPT2 (Radford et al., 2019) and T5 (Raffel et al., 2020), respectively, as backbones for both code understanding and generation. Different learning methods including learning from examples (Ellis et al., 2021) and neural-symbolic methods Nye et al. (2020) were explored. Recently, AlphaCode (Li et al., 2022) combined large transformer models pre-trained on massive program data with large-scale model sampling, showing competitive performance in programming competitions. All these models

mainly focused on training more power code-generation models and simply use sampling (Fan et al., 2018) or beam-search (Graves, 2012) during the Transformers' generation process.

Test cases for program synthesis. Our work is also related to using unit tests (Myers, 1979) for program synthesis. Tufano et al. (2020a;b) propose to generate test cases and corresponding accurate assert statements with Transformer models (Vaswani et al., 2017a). Recently, Roziere et al. (2022) leverage automated unit tests to construct parallel training data for unsupervised code translation. Chen et al. (2018); Gupta et al. (2020); Chen et al. (2021b) directly synthesize domain-specific programs from input-output pairs without problem description. Ellis et al. (2019) use test cases to train a reinforcement learning agent and use sequential Monte-Carlo sampling method for code generation. Unlike the prior work, we use unit-testing results as reward signals for the Monte-Carlo tree search (MCTS) planning algorithm, which is further integrated with a Transformer-based large language model to generate better codes.

Planning and reinforcement learning for code generation. The code generation problem has been formulated as a sequential decision making problem (Bunel et al., 2018; Chen et al., 2018), which enables designing and applying reinforcement learning (RL) and planning algorithms for code generation. RL has been used for both the training phase and the decoding phase of the Transformer for code generation. In the training phase, Le et al. (2022); Xu et al. (2019b) use an RL objective that optimizes the correctness of the generated programs instead of optimizing their similarity to reference solutions. In the decoding phase, the Monte-Carlo tree search (MCTS) algorithm has been applied to search for high-quality codes. However, MCTS is unable to scale to larger domains and is only used to generate domain-specific languages or for restricted programming synthesis tasks like assembly code generation (Xu et al., 2019a), Java bytecode translation (Lim & Yoo, 2016), and robot planning (Matulewicz, 2022). These tasks have a much smaller scale than generating Python codes in our work. Therefore, their framework does not require a large language model and does not need to address the challenge of incorporating a planning algorithm with a large language model.

MCTS is more efficient and applicable to large domains when combined with deep learning or with a default policy as prior knowledge (Gelly & Silver, 2011; Silver et al., 2016; Simmons-Edler et al., 2018). We consider the same overall recipe in our work. Specifically, we contributed to integrating the large language model with MCTS to design a novel algorithm that is capable of solving competitive programming problems, and designed mechanisms to improve its efficiency.

Planning in natural language processing. Planning algorithms like MCTS have also been used to find the optimal text outputs for different natural language processing (NLP) tasks. For example, Scialom et al. (2021); Leblond et al. (2021); Chaffin et al. (2022) use pre-trained discriminators or pre-defined metrics as reward functions. We want to emphasize that we are the first to combine MCTS with large language models for general-purpose programming language generation. We design the interfaces between these two components and deal with the unique challenges of making the framework computationally efficient. Concurrently, other search algorithms like A* algorithm (Hart et al., 1968) are also applied in the Transformer decoding process. Lu et al. (2021) consider the constrained text generation problem and integrate A* with the beam search algorithm. However, their constraints are expressed as a formal logic expression, which is different from maximizing the pass rate value in our problem.

3 METHOD

3.1 OVERVIEW

We consider the code generation problem for competitive programming, illustrated in Fig. 1. An agent is given the problem description of a coding problem. It requires the agent to understand the problem from the natural language description and generate a program that solves the problem. Similar to Li et al. (2022); Chen et al. (2018); Ellis et al. (2019), we assume that the agent has access to a set of test cases, where a test case is a pair of input, output strings. Given the input string, the agent is expected to generate a program that produces an output that exactly matches the test case's output string. The objective is to generate a program that passes the most number of test cases. To determine if the agent is able to generate programs that generalize to unseen test cases, we divide the test cases into *public* test cases and *private* test cases, following the terms in Li et al. (2022). The agent can only access the public test cases during the program generation process, while we use the private test cases to evaluate the programs it generates.

Problem Statement

Given is a string S. Replace every character in S with x and print the result.

Constraints

(1). S is a string consisting of lowercase English letters.

(2). The length of S is between 1 and 100 (inclusive).

Input

Input is given from Standard Input in the following format: S

Output

Replace every character in S with x and print the result.

Sample Test Input

$sardine$

Sample Test Output

$xxxxxxx$

```
s=input()
s=list(s)
for i in range(len(s)):
    for j in range(len(s)):
        if s[i]=="x":
            s[i]=j
    print("".join(s))

```

Beam Search (Pass Rate: 0.00).

```
s=input()
s=list(s)
for i in range(len(s)):
    if s[i]=="x":
        s[i]="x"
    else:
        continue
print("".join(s))
```

Sampling + Filtering (Pass Rate: 0.22).

```
s=str(input())
for i in range(len(s)):
    if s[i]!="x":
        s=s[:i]+"x"+s[i+1:]

print(s)

```

PG-TD (Pass Rate: 1.00).

Figure 1: A code generation example for competitive programming, with the problem description (top) and the programs generated by baseline algorithms and our PG-TD algorithm (bottom).

Transformer models (Li et al., 2022; Hendrycks et al., 2021) have been widely applied for code generation thanks to its capacity in sequence-to-sequence modeling. In the Transformer's generation process, beam search (Graves, 2012) and sampling (Fan et al., 2018; Dabre & Fujita, 2020) are adopted to generate code sequences. However, these algorithm cannot easily optimize an objective different from what it is trained on (the similarity to the reference solutions). So we cannot directly use these generation algorithms to generate programs aiming to pass the most test cases.

On the other hand, a planning algorithm can directly optimize the pass rate or any desirable programming objective. To use a planning algorithm, we follow Bunel et al. (2018); Ellis et al. (2019) to first formulate the code generation problem as a Markov decision process (MDP) (Sutton & Barto, 2018). A **state** s is the concatenation of the problem description and a partial or complete program, where a complete program ends with a special terminal token. An **action** a is a token in the vocabulary set of the Transformer, which contains the termination action (the terminal token) that indicates that the agent believes the program is complete. The **transition function** deterministically concatenates the state s with the token a, and an episode ends when the agent takes the termination action. The **reward** of state s is the pass rate of the program on the public test cases when s is a complete program (*i.e.,* when the last token of s is the terminal token). The reward of a partial program is always 0.

There is rich literature on finding the optimal policy for an MDP. In this paper, we consider a tree search-based planning algorithm inspired by Monte-Carlo tree search (MCTS), illustrated in Fig. 2. The algorithm maintains a tree structure where nodes correspond to states and edges correspond to actions. The algorithm starts from the the root node (the initial state) and searches the state space to find terminal states with high rewards. It maintains 1) the number of times each node is visited and 2) a value function for each edge s, a, which is the maximum value obtained by starting in s and taking action a. The algorithm would visit nodes with either higher Q values (as they lead to higher-quality programs) or with smaller visit numbers (as they are under-explored). In the following part of this section, we describe how we integrate the MCTS algorithm in the generation process of a Transformer.

3.2 Planning-Guided Transformer Decoding

Now we are ready to answer the question we asked in the introduction: Can we integrate a planning algorithm with a pretrained code generation Transformer to generate better programs? We design a

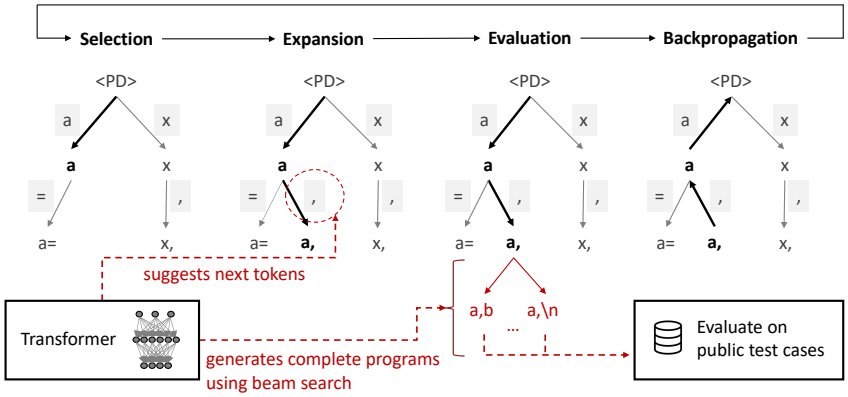


Figure 2: Illustration of using the Monte Carlo tree search algorithm in the Transformer generation process for code generation. <PD> stands for problem description.

Transformer generation algorithm where MCTS is called iteratively to perform lookahead planning and return with the token that will lead to high-quality programs. The MCTS algorithm alone may not be able to find high-quality codes due to the large search space. So the conventional Transformer beam search algorithm and the next-token probabilities are used within the MCTS algorithm to guide the search process.

The main generation process is described in Algorithm 2. In each step, PG-TD is called to generate a token a_t. We provide the pseudocode of PG-TD in Algorithm 1 and illustrate the whole process in Figure 2. The PG-TD algorithm follows the same framework as the standard MCTS algorithm. We specify k_{max} to be the maximum number of children that any node may have. As our problem has a huge action space (the set of all tokens), we specifically constrain the number of children so that the agent would not explore too many actions of the same state.

For completeness, we provide all the steps in the pseudocode and in Sec. C.1. Here, we focus on the expansion and the evaluation steps where the Transformer is used to provide heuristics for the search process. The Transformer provides two functions for PG-TD: GENERATE_SEQ(s, b) generates a list of b programs using the Transformer beam search algorithm with the prefix s, where b is the beam size; TOP_K(s, k) returns the k-th most likely next token starting from s.

In the expansion step, after a node in the tree is selected, we add a new node to its children which corresponds to a possible next state (for succinctness, a node also refers to the state that it represents). Adding a random token as in standard MCTS may very likely cause a syntax error. So we call the TOP_K function to get the most likely next tokens (or actions). When the current

Algorithm 1 The PG-TD algorithm.

Require: $root$: the current state; $rollouts$: the number of MCTS rollouts; k_{max}: the maximum number of children of any node; b: the number of beams for Transformer beam search

for $i \leftarrow 1, 2, \ldots, rollouts$ **do**
  $node \leftarrow root$
  # Selection
  while $|node.children| = k_{max}$ **do**
    $node \leftarrow$ UCB_SELECT$(node.children)$
  end while
  # Expansion
  $children\_num \leftarrow |node.children|$
  $act \leftarrow$ TOP_K$(node, k = children_num + 1)$
  $new\_node \leftarrow$ the state reached by taking action act in $node$
  Add new_node to $node.children$
  # Evaluation
  $program\_list \leftarrow$ GENERATE_SEQ(new_node, b)
  $r \leftarrow \max_{p \in program\_list}$ GET_REWARD(p)
  # Backpropagation
  Update and the values of new_node and its ancestors in the tree with r
end for
return the action with the highest value starting from $root$

Algorithm 2 The main function.

$s_0 \leftarrow$ <PD> # program description
$t \leftarrow 0$
while $s_t[-1] \neq$ <|endoftext|> and $t < max\_length$ **do**
  $a_t \leftarrow$ PG-TD$(s_t, rollouts, k_{max}, b)$
  $s_{t+1} \leftarrow$ CONCATENATE(s_t, a_t)
  $t \leftarrow t + 1$
end while
return s_t

node already has $children\_num$ children, we get the $(children\_num+1)$-st most likely next token, and add the corresponding next state, $next\_node$, to the children of $node$ (Line 9-11).

In the evaluation step, we need to evaluate the quality of new_node. Note that new_node may still be a partial program. We cannot directly evaluate the quality of a partial program as we do not know how it will be completed, and how many test cases it will pass. Here, we use the Transformer again by calling the GENERATE_SEQ function to generate complete programs from the current node using the beam search algorithm. We set the size of beam search, b, to indicate how many complete programs that we want to generate from the current partial program. We run these generated programs on the public test cases to get their rewards, and set the value of new_node to be the highest reward of the complete programs (Line 13-14).

3.3 Information Sharing Between Transformer and MCTS

A keen reader may notice that if we follow the algorithm described above, there may be a lot of repeated computations. The key observation is that the Transformer beam search algorithm also *implicitly builds a tree structure*, which can be used by the future iterations of MCTS. In this section, we describe how we improve the algorithm's efficiency by sharing information in the Transformer beam search algorithm with MCTS.

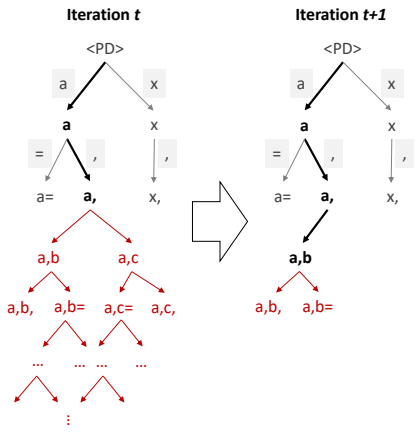


Figure 3: Illustration for caching in the PG-TD algorithm. The MCTS part is visualized in black color and the Transformer beam search part is in red color.

Consider the example in Fig. 3, which shows two iterations of PG-TD. In the evaluation step of the t-th iteration, the Transformer beam search algorithm implicitly builds a tree to find the most likely sequences within a beam, visualized in red color in Fig. 3 (left). Because we only keep b partial programs in the beam, it is a tree where only b nodes with the highest likelihood are expanded at each level (in the illustration, $b = 2$). Other nodes are dropped and no longer considered by the beam search algorithm. In the $(t + 1)$-st iteration, if MCTS selects "`a,`" , such a state is already visited in the Transformer beam search in the t-th iteration. When the algorithm needs to find the top-k most-likely next tokens, such information is already obtained in the t-th iteration and can be reused without recomputation. In our implementation, we cache the tree structure generated by the Transformer beam search algorithm. We call this implementation **tree structure caching**.

We can also cache the complete programs generated during the PG-TD evaluation step. In the evaluation step of the t-th iteration (Fig. 3), suppose the greedily-optimal sequence is "`a,b=...`". In the $(t + 1)$-st iteration, to generate a sequence starting with "`a,b`", the Transformer beam search algorithm will generate the same sequence "`a,b=...`" as before. (This is not necessarily true when the beam size is larger than 1. We will clarify this in Sec. C.2.) To improve the efficiency of GENERATE_SEQ, we cache the sequences that have been generated during the evaluation step. In the evaluation step of future iterations, PG-TD will check if the current state matches the prefix of any sequence that has been generated before, and use the generated sequence directly without calling the Transformer beam search function. We call this implementation **sequence caching**.

One may be concerned about the space efficiency of tree structure caching and sequence caching, as they keep saving trees and sequences for each node in the tree. Fortunately, we only need to keep the caches that are consistent with the current state. Once a new action is taken, we maintain the cached data that are useful for the new state. We will provide more details on the caching methods in Sec. C.2.

4 Empirical Evaluation

In this section, we empirically examine the effectiveness and efficiency of our PG-TD algorithm by answering the following questions. **Q1:** Does PG-TD generate better programs than using the Transformer beam search algorithm and other competing baseline methods? Is our algorithm model-agnostic, showing improvement when applied to different Transformer models? **Q2:** Is the

Models		Pass Rate (%)		Strict Accuracy (%)	
		APPS Intro.	CodeContests	APPS Intro.	CodeContests
APPS GPT-2 (1.5B)	Beam Search	11.95	5.10	5.50	0.00
	Sampling + Filtering	16.84	16.94	7.30	1.81
	SMCG-TD	17.89	13.97	8.10	**3.03**
	PG-TD	**23.14**	**20.85**	**11.00**	1.81
APPS GPT-Neo (2.7B)	Beam Search	14.32	5.73	6.70	0.00
	Sampling + Filtering	20.79	18.19	10.10	1.82
	SMCG-TD	21.00	18.08	10.40	1.21
	PG-TD	**26.14**	**21.08**	**13.20**	**2.42**

Table 1: Average pass rates of PG-TD and baseline methods on competitive programming datasets.

tree search algorithm an effective planning algorithm in PG-TD? Is it better than other planning or sampling-based algorithms? **Q3:** Is PG-TD efficient in terms of the number of times it runs the Transformer beam search algorithm and also the computation time it consumes? **Q4:** Are the caching methods effective in saving the computation time? **Q5:** Can we use the samples generated by PG-TD in its search process and finetune the Transformer to generate even better solutions?

Datasets and models. Recently, several competitive programming datasets have been made available for benchmarking code generation algorithms. For each programming problem, they include the program description in natural language, sample solutions, and test cases. We evaluated PG-TD and the baseline algorithms on some popular benchmark datasets: APPS (Hendrycks et al., 2021) and CodeContests in AlphaCode (Li et al., 2022). APPS does not split public vs. private test cases. So we split all the test cases of a program evenly into two sets, where the first set is used as the public test cases for the algorithms to optimize the pass rate, and the second set is used as the private test cases for evaluating the generated programs. For CodeContests, we use their public and generated test cases as our public test cases, and their private test cases as our private test cases. To show that PG-TD is model-agnostic and can be applied to different pre-trained Transformers and achieve a better performance, we use two popular pre-trained code-generation Transformers in the literature: GPT-2, GPT-Neo that are finetuned on the APPS training dataset (Hendrycks et al., 2021).

Algorithms. We compare PG-TD with the following algorithms. **Beam search** only uses the Transformer beam search algorithm to generate the whole program, without using the test cases. This is the method used in Hendrycks et al. (2021); Chen et al. (2021a). We use the beam size of 5 in the Transformer generation function, which is the same design choice as in Hendrycks et al. (2021).

We also implemented two baselines that use Transformer to sample programs and filter them using the test cases. **Sampling + Filtering (S+F)** generates a set of programs using the Transformer sampling algorithm. Once the programs are all generated, it computes the pass rates of all the programs on the public test cases and returns the program with the highest pass rate. To avoid generating low-probability tokens that may fail the program completely, we use top-3 sampling, that is, the Transformer only samples the token that is in the top-3 most-likely tokens in each step. This method is similar to the algorithm in AlphaCode (Li et al., 2022), except that we did not perform clustering as we generate a much smaller number of programs than them. To determine the effectiveness of the tree search algorithm, we considered another baseline, **Sequential Monte-Carlo-Guided Transformer Decoding (SMCG-TD)**, that replaces the the tree search algorithm component with a sequential Monte-Carlo algorithm (Ellis et al., 2019). It is an iterative algorithm that maintains a population of partial programs. It determines the *fitness* of a partial program using the Transformer in the same way as the evaluation step in PG-TD. Partial programs with higher fitness scores are more likely to be selected in the next iteration. We leave more details about the baseline algorithms in Sec C.3.

For PG-TD, we set the parameters $k_{max} = 3$, $rollouts = 3$, $b = 1$. For the baseline methods, we sample 512 programs in S+F, and maintain a set of 10 partial programs in each iteration for SMCG-TD. We will use these parameters for the following experiments unless noted otherwise. The experiments are conducted on the APPS introductory dataset (1000 problems) (Hendrycks et al., 2021) and CodeContests (165 problems) (Li et al., 2022). We use the same metrics as in Hendrycks et al. (2021), pass rates and strict accuracies of the private test cases. The private test cases are hidden from the agent during the code generation process. The pass rate is the average percentage

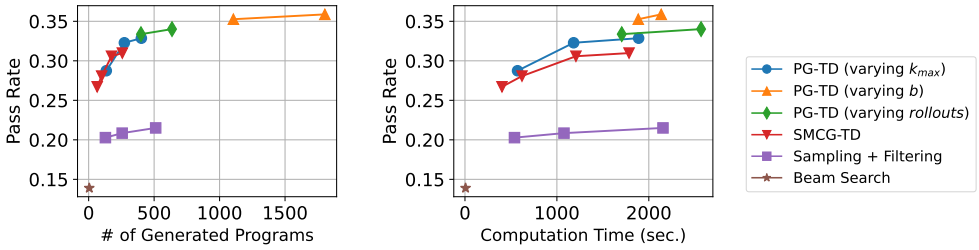


Figure 4: Pass rates of PG-TD and baseline algorithms vs. the number of programs generated using the Transformer (left) and the computation time (middle) on the first 500 problems in the APPS introductory dataset, using the APPS GPT-2 Transformer model.

Method	Time (sec.)
With both caching methods	**1312.27**
W/ only sequence caching	1430.63
W/ only tree structure caching	1899.17
W/o caching	2206.38

Table 2: Affects of using caching for PG-TD on the first 100 problems on the APPS introductory dataset, using the APPS GPT-2 (1.5B).

Method	Pass Rate (%)	Strict Acc. (%)
Original	14.32	6.70
FT w/ S+F	15.24	**7.90**
FT w/ PG-TD	**16.54**	**7.90**

Table 3: Performance on APPS introductory dataset with finetuned (FT) APPS GPT-2 (2.7B), using Beam Search for code generation.

of the private test cases that the generated programs pass. The strict accuracy is the percentage of the problems where the generated programs pass all the private test cases.

Effectiveness of PG-TD. As shown in Table. 1, PG-TD consistently outperforms all other baselines in all the settings under the pass rate metric. As we optimize the pass rate, we expect our algorithm to outperform other baselines under strict accuracy as well. This is almost true except that using the GPT-2 (2.7B) model on the CodeContests dataset, our algorithm is slightly outperformed by SMCG-TD. Overall, these results confirm that our algorithm indeed generates programs that pass more test cases than the baseline methods (answering **Q1**). Specifically, S+F and SMCG-TD both use test cases to filter programs generated by the Transformer, while their performance is overall outperformed by PG-TD. So the tree search-based planning algorithm is indeed powerful enough and can be effectively integrated with Transformer (answering **Q2**). We also report the performance of PG-TD and S+F under the n@k and pass@k metrics (Li et al., 2022) in Sec. A.

Efficiency of PG-TD. To show PG-TD's efficiency, we report the pass rates vs. the number of generated programs by Transformer and the computation time on a smaller dataset in Fig. 4. We set $k_{max} = 3$, $rollouts = 3$, and $b = 1$ as the default setting and only changes one parameter at a time. We vary these parameters to be $k_{max} = 2, 3, 4$, $rollouts = 6, 9$, and $b = 3, 5$. As we expect, increasing k_{max}, $rollouts$, and b all helps PG-TD achieve higher pass rates, while increasing b helps improve the pass rate the most as the Transformer beam search algorithm generates more high-likelihood candidates.

We also consider different sample sizes for S+F. We choose sample sizes of 128, 256, and 512. With the same number of samples, S+F generates programs with much lower pass rates (answering **Q3**). This confirms our observation that S+F, like AlphaCode (Li et al., 2022), generates samples only according to the Transformer's generation probabilities, without taking their pass rates into consideration *until the programs are all generated*. PG-TD actively considers the pass rates of the generated programs during the generation process, which achieves a higher efficiency. For SMCG-TD, we set the size of population to be 10, 20, 50, and 100. Since SMCG-TD also uses the Transformer and public test cases to guide the generation process, we expect that its performance is close to the default setting of PG-TD with a larger size of population. However, due to its sampling nature, it cannot do multi-step lookahead search as in PG-TD. So SMCG-TD's performance is worse than the variations of PG-TD. More details are discussed in Sec. C.3.

Effectiveness of caching. In terms of the design choice of tree structure caching and sequence caching, we performed ablative studies to verify their efficiencies. In Table 2, we compare versions of PG-TD with and without tree structure caching and sequence caching. As we expect, without sequence caching, the algorithm needs to regenerate whole sequences, ending up with consuming

```
a, b = map(int, input().split())
print(a + b)
```

Default reward function.

```
#!/usrbin/env python3
# coding =utf-8 #py3 runs on py2
a,b=map(int,input().split())
print(a+b)
```

Reward function with comment encouragement.

Figure 5: A code generation example with code comment encouragement. More details about the problem description and implementation details can be found in the Sec. B of the appendix.

much more time. Without tree structure caching, the algorithm is slightly slower as it needs to call Transformer to get the most-likely next tokens (answering **Q4**).

Finetuning transformer with PG-TD-generated samples. Since we are able to generate solutions with high pass rates using our PG-TD algorithm, can we use these generated solutions to finetune the code generation Transformers and further improve their performance? This may effectively solve the problem that high-quality human-written programs that can be used to train the code generation Transformers are scarcely available. Concretely, we first run PG-TD on the training set of APPS. We then add the generated solutions with pass rates larger than 80% to the APPS sample solution set, and use the augmented solution set to finetune the GPT-2 (2.7B) model. With the finetuned Transformoer model, we run beam search to generate solutions on the test set to see if it has a better performance. We use the first 500 problems in the APPS interview-level test set for validation and the APPS introductory-level test set for testing. The results are reported in Table 3. After finetuning with PG-TD-generated solutions, we see improvement in both pass rate and strict accuracy. More details are in Sec. C.4.

Optimizing other code generation objectives. Beyond the pass rate, we can make the algorithm versatile by using different reward functions. We consider two new objectives, **code length penalty** and **comment encouragement**. As shown in Table 4, the code length penalty reward function makes the model generate more concise codes; the comment encouragement reward function encourages models to generate codes with more comments. Both objectives still achieve reasonable pass rates on the public test cases. We provide a qualitative example for comment encouragement in Figure 5 and more details of the reward functions in Sec. B of the appendix.

Methods	Code length ↓	Comment number ↑	Pass rate ↑
Default	248.42	0.68	**23.14**
Length Penalty	**190.73**	-	22.82
Comment Encouragement	-	**3.11**	21.65

Table 4: Performance of controllable code generation. Code length denotes the length of the generated code string and comment number denotes the number of code lines that contains comments.

5 Discussion and Conclusion

In summary, we proposed a novel algorithm that uses the powers of a pre-trained Transformer and the Monte-Carlo tree search algorithm. We evaluate our algorithm and empirically show that our algorithm generates programs of higher quality compared with competing baseline algorithms in different settings. We also design model structure-specific caching mechanisms which contribute to saving computational expenses. We show that our algorithm is versatile and can generate codes under objectives other than the pass rate without finetuning the Transformer. We hope our work can inspire more ways of integrating planning algorithms into the Transformer generation process for code generation or other problems.

One limitation of our algorithm is its reliance on test cases, although we find that even a small number of test cases can help find better solutions (Sec. A, Table 7). Our algorithm is also more computationally expensive than the pure Transformer beam search algorithm since ours needs to run the beam search algorithm for multiple times within MCTS. In the future, we consider improving the framework's performance by using a value function to estimate the programs' pass rates, similar to the evaluation function in Silver et al. (2016) for mastering Go. We can also learn a neural state representation for partial programs as in Chen et al. (2018). We also consider using parallel MCTS (Chaslot et al., 2008) for the evaluation step to parallelize the computation for different states.

REFERENCES

Wasi Ahmad, Saikat Chakraborty, Baishakhi Ray, and Kai-Wei Chang. Unified pre-training for program understanding and generation. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 2655–2668, Online, June 2021. Association for Computational Linguistics. 1

Jacob Austin, Augustus Odena, Maxwell Nye, Maarten Bosma, Henryk Michalewski, David Dohan, Ellen Jiang, Carrie Cai, Michael Terry, Quoc Le, and Charles Sutton. Program Synthesis with Large Language Models. *arXiv:2108.07732 [cs]*, August 2021. URL http://arxiv.org/abs/2108.07732. arXiv: 2108.07732. 2

Rudy Bunel, Matthew Hausknecht, Jacob Devlin, Rishabh Singh, and Pushmeet Kohli. Leveraging Grammar and Reinforcement Learning for Neural Program Synthesis. *ICLR*, May 2018. URL http://arxiv.org/abs/1805.04276. arXiv: 1805.04276. 3, 4

Antoine Chaffin, Vincent Claveau, and Ewa Kijak. Ppl-mcts: Constrained textual generation through discriminator-guided mcts decoding. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 2953–2967, 2022. 3

Guillaume MJ-B Chaslot, Mark HM Winands, and HJVD Herik. Parallel monte-carlo tree search. In *International Conference on Computers and Games*, pp. 60–71. Springer, 2008. 9

Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde de Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, Alex Ray, Raul Puri, Gretchen Krueger, Michael Petrov, Heidy Khlaaf, Girish Sastry, Pamela Mishkin, Brooke Chan, Scott Gray, Nick Ryder, Mikhail Pavlov, Alethea Power, Lukasz Kaiser, Mohammad Bavarian, Clemens Winter, Philippe Tillet, Felipe Petroski Such, Dave Cummings, Matthias Plappert, Fotios Chantzis, Elizabeth Barnes, Ariel Herbert-Voss, William Hebgen Guss, Alex Nichol, Alex Paino, Nikolas Tezak, Jie Tang, Igor Babuschkin, Suchir Balaji, Shantanu Jain, William Saunders, Christopher Hesse, Andrew N. Carr, Jan Leike, Josh Achiam, Vedant Misra, Evan Morikawa, Alec Radford, Matthew Knight, Miles Brundage, Mira Murati, Katie Mayer, Peter Welinder, Bob McGrew, Dario Amodei, Sam McCandlish, Ilya Sutskever, and Wojciech Zaremba. Evaluating Large Language Models Trained on Code. *arXiv:2107.03374 [cs]*, July 2021a. URL http://arxiv.org/abs/2107.03374. arXiv: 2107.03374. 1, 2, 7, 15

Xinyun Chen, Chang Liu, and Dawn Song. Execution-guided neural program synthesis. In *International Conference on Learning Representations*, 2018. 3, 9

Xinyun Chen, Dawn Song, and Yuandong Tian. Latent execution for neural program synthesis beyond domain-specific languages. *Advances in Neural Information Processing Systems*, 2021b. 3

Raj Dabre and Atsushi Fujita. Softmax tempering for training neural machine translation models. *arXiv*, 2020. 1, 4

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1423. URL https://aclanthology.org/N19-1423. 2

Kevin Ellis, Catherine Wong, Maxwell Nye, Mathias Sablé-Meyer, Lucas Morales, Luke Hewitt, Luc Cary, Armando Solar-Lezama, and Joshua B Tenenbaum. Dreamcoder: Bootstrapping inductive program synthesis with wake-sleep library learning. In *Proceedings of the 42nd ACM SIGPLAN International Conference on Programming Language Design and Implementation*, pp. 835–850, 2021. 2

Kevin M. Ellis, Maxwell Nye, Yewen Pu, Felix Sosa, Joshua Tenenbaum, and Armando Solar-Lezama. Write, execute, assess: Program synthesis with a repl. *Neural Information Processing Systems*, 2019. 3, 4, 7

Angela Fan, Mike Lewis, and Yann Dauphin. Hierarchical neural story generation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 889–898, 2018. 1, 3, 4

Zhangyin Feng, Daya Guo, Duyu Tang, Nan Duan, Xiaocheng Feng, Ming Gong, Linjun Shou, Bing Qin, Ting Liu, Daxin Jiang, et al. Codebert: A pre-trained model for programming and natural languages. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pp. 1536–1547, 2020. 2

Sylvain Gelly and David Silver. Monte-carlo tree search and rapid action value estimation in computer go. *Artificial Intelligence*, 175(11):1856–1875, 2011. 3

Alex Graves. Sequence transduction with recurrent neural networks. In *International Conference on Machine Learning: Representation Learning Workshop*, 2012. 3, 4

Daya Guo, Shuo Ren, Shuai Lu, Zhangyin Feng, Duyu Tang, LIU Shujie, Long Zhou, Nan Duan, Alexey Svyatkovskiy, Shengyu Fu, et al. Graphcodebert: Pre-training code representations with data flow. In *International Conference on Learning Representations*, 2020. 2

Kavi Gupta, Peter Ebert Christensen, Xinyun Chen, and Dawn Song. Synthesize, execute and debug: Learning to repair for neural program synthesis. *Advances in Neural Information Processing Systems*, 2020. 3

Peter E Hart, Nils J Nilsson, and Bertram Raphael. A formal basis for the heuristic determination of minimum cost paths. *IEEE transactions on Systems Science and Cybernetics*, 4(2):100–107, 1968. 3

Dan Hendrycks, Steven Basart, Saurav Kadavath, Mantas Mazeika, Akul Arora, Ethan Guo, Collin Burns, Samir Puranik, Horace He, Dawn Song, and Jacob Steinhardt. Measuring coding challenge competence with APPS. *NeurIPS*, 2021. 1, 4, 7, 14, 16

Aditya Kanade, Petros Maniatis, Gogul Balakrishnan, and Kensen Shi. Learning and evaluating contextual embedding of source code. In *International Conference on Machine Learning*, pp. 5110–5121. PMLR, 2020. 2

Levente Kocsis and Csaba Szepesvári. Bandit based monte-carlo planning. In *European conference on machine learning*, pp. 282–293. Springer, 2006. 20

Hung Le, Yue Wang, Akhilesh Deepak Gotmare, Silvio Savarese, and Steven C. H. Hoi. CodeRL: Mastering Code Generation through Pretrained Models and Deep Reinforcement Learning. arXiv, July 2022. doi: 10.48550/arXiv.2207.01780. URL http://arxiv.org/abs/2207.01780. arXiv:2207.01780 [cs]. 3, 23

Rémi Leblond, Jean-Baptiste Alayrac, Laurent Sifre, Miruna Pislar, Jean-Baptiste Lespiau, Ioannis Antonoglou, Karen Simonyan, and Oriol Vinyals. Machine translation decoding beyond beam search. *arXiv preprint arXiv:2104.05336*, 2021. 3

Yujia Li, David Choi, Junyoung Chung, Nate Kushman, Julian Schrittwieser, Rémi Leblond, Tom Eccles, James Keeling, Felix Gimeno, and Agustin Dal Lago. Competition-Level Code Generation with AlphaCode. *arXiv preprint arXiv:2203.07814*, 2022. 1, 2, 3, 4, 7, 8, 14, 16

Jinsuk Lim and Shin Yoo. Field report: Applying monte carlo tree search for program synthesis. In *International Symposium on Search Based Software Engineering*, pp. 304–310. Springer, 2016. 3

Ximing Lu, Sean Welleck, Peter West, Liwei Jiang, Jungo Kasai, Daniel Khashabi, Ronan Le Bras, Lianhui Qin, Youngjae Yu, Rowan Zellers, Noah A. Smith, and Yejin Choi. NeuroLogic A*esque Decoding: Constrained Text Generation with Lookahead Heuristics, December 2021. URL http://arxiv.org/abs/2112.08726. arXiv:2112.08726 [cs]. 3

Nadia Matulewicz. Inductive program synthesis through using monte carlo tree search guided by a heuristic-based loss function. 2022. 3

Glenford J Myers. The art of software testing. 1979. 3

Maxwell Nye, Armando Solar-Lezama, Josh Tenenbaum, and Brenden M Lake. Learning compositional rules via neural program synthesis. *Advances in Neural Information Processing Systems*, 33:10832–10842, 2020. 2

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019. 2, 32

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 2020. URL http://jmlr.org/papers/v21/20-074.html. 2

Baptiste Roziere, Marie-Anne Lachaux, Lowik Chanussot, and Guillaume Lample. Unsupervised translation of programming languages. *Advances in Neural Information Processing Systems*, 33, 2020. 2

Baptiste Roziere, Jie M Zhang, Francois Charton, Mark Harman, Gabriel Synnaeve, and Guillaume Lample. Leveraging automated unit tests for unsupervised code translation. In *International Conference on Learning Representations*, 2022. 1, 3, 16

Thomas Scialom, Paul-Alexis Dray, Jacopo Staiano, Sylvain Lamprier, and Benjamin Piwowarski. To beam or not to beam: That is a question of cooperation for language gans. *Advances in neural information processing systems*, 34:26585–26597, 2021. 3

David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, and Marc Lanctot. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484, 2016. Publisher: Nature Publishing Group. 3, 9

David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, and Adrian Bolton. Mastering the game of go without human knowledge. *nature*, 550(7676):354–359, 2017. Publisher: Nature Publishing Group. 21

Riley Simmons-Edler, Anders Miltner, and Sebastian Seung. Program synthesis through reinforcement learning guided tree search. *arXiv preprint arXiv:1806.02932*, 2018. 3

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, October 2018. ISBN 978-0-262-35270-3. 4

Michele Tufano, Dawn Drain, Alexey Svyatkovskiy, Shao Kun Deng, and Neel Sundaresan. Unit test case generation with transformers and focal context. *arXiv*, 2020a. 3

Michele Tufano, Dawn Drain, Alexey Svyatkovskiy, and Neel Sundaresan. Generating accurate assert statements for unit test cases using pretrained transformers. *arXiv*, 2020b. 3

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 2017a. 1, 3

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need. *arXiv:1706.03762 [cs]*, December 2017b. URL http://arxiv.org/abs/1706.03762. arXiv: 1706.03762. 1

Yue Wang, Weishi Wang, Shafiq Joty, and Steven C.H. Hoi. CodeT5: Identifier-aware Unified Pre-trained Encoder-Decoder Models for Code Understanding and Generation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 8696–8708, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.emnlp-main.685. URL https://aclanthology.org/2021.emnlp-main.685. 1, 2

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, and Morgan Funtowicz. Huggingface's transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*, 2019. 16

Yifan Xu, Lu Dai, Udaikaran Singh, Kening Zhang, and Zhuowen Tu. Neural program synthesis by self-learning. *arXiv preprint arXiv:1910.05865*, 2019a. 3

Yifan Xu, Lu Dai, Udaikaran Singh, Kening Zhang, and Zhuowen Tu. Neural Program Synthesis By Self-Learning. October 2019b. doi: 10.48550/arXiv.1910.05865. URL https://arxiv.org/abs/1910.05865v1. 3

APPENDIX

In this appendix, we supplement the main paper by providing more thorough empirical evaluations to back up our claims and more detailed descriptions of the algorithms to help readers better understand our paper.

This appendix is organized as follows.

- In Sec. A, we provide empirical results on more competitive coding datasets and more comprehensive results of our algorithm and the baseline algorithms. We also include the license information of the datasets we use.
- In Sec. B, we provide empirical evidence for our claims in the discussion section that our algorithm is versatile and can be used to optimize different code generation objectives other than optimizing the pass rates. We consider the objectives of code length penalty and comment encouragement.
- In Sec. C, we provide more details on the components in PG-TD as well as the baseline algorithms.
- In Sec. D, we illustrate more examples on the codes generated by our PG-TD algorithm and the baseline algorithms, and also provide a step-by-step visualization of execution of PG-TD.
- In Sec. E, we discuss more on the advantages and the potential negative social impacts of our algorithm.

A EMPIRICAL EVALUATION

We report the performance of our algorithm and other baselines on the whole APPS dataset (Hendrycks et al., 2021) and CodeContests (Li et al., 2022). The APPS dataset contains coding problems at three levels: introductory (1000 problems), interview (3000 problems), and competition (1000 problems). Our experiments are run on machines with two Intel(R) Xeon(R) Gold 6258R CPUs (@ 2.70GHz), and one V100-SXM2 GPU.

In Table 5, we can see that our algorithm consistently has the highest pass rate on all the datasets, which further supports **Q1** in the main paper.

		Pass Rate (%)				Strict Accuracy (%)			
		APPS Intro.	APPS Inter.	APPS Comp.	CodeContests	APPS Intro.	APPS Inter.	APPS Comp.	CodeContests
GPT-2	Beam Search	11.95	9.55	5.04	5.10	5.50	2.10	1.00	0.00
	S+F	16.84	17.34	9.16	16.94	7.30	3.56	2.00	1.81
	SMCG-TD	17.89	17.85	9.94	13.97	8.10	3.20	2.10	**3.03**
	PG-TD	**23.14**	**24.38**	**11.49**	**20.85**	**11.00**	**5.36**	**2.70**	1.81
GPT-Neo	Beam Search	14.32	9.80	6.39	5.73	6.70	2.00	2.10	0.00
	S+F	20.79	18.03	10.12	18.19	10.10	3.13	2.20	1.82
	SMCG-TD	21.00	18.54	9.09	18.08	10.4	3.63	2.20	1.21
	PG-TD	**25.64**	**22.13**	**11.09**	**21.08**	**13.20**	**4.83**	**2.50**	**2.42**

Table 5: Average percentages of pass rates and strict accuracies of PG-TD and baseline methods on competitive programming datasets.

In Fig. 6, we supplement Fig. 4 by reporting the strict accuracies of PG-TD and baseline algorithms. They have the similar patterns as the pass rate results. However, since we only optimize the pass rates, the performance of each algorithm does not monotonically increase by increasing the size of the tree in PG-TD or increasing the sample sizes for Sampling and SMCG-TD.

n@k and pass@k metrics. We also report the performance of PG-TD and Sampling under the n@k and pass@k metrics (Li et al., 2022). Given a problem, each algorithm is asked to generate k programs and submit n programs for evaluation (they submit the n programs with the highest reward on the public test cases). n@k is the proportion of the problems where any of the n submitted programs passes all the private test cases. pass@k is the proportion of the problems where any of the k generated programs passes all the private test cases (effectively k@k). For PG-TD, the k samples are the complete programs found in the first k rollouts. The results are reported in Table 6, evaluated on the first 500 APPS introductory problems. We observe that with the same number of generated samples (k), PG-TD finds programs that are more likely to pass all the private test cases.

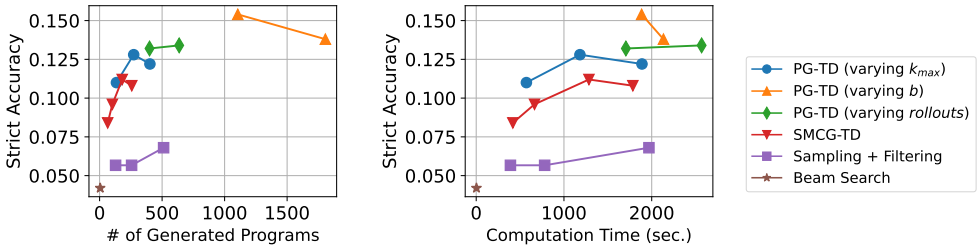


Figure 6: Strict accuracies of PG-TD and baseline algorithms vs. the number of programs generated using the Transformer (left) and the computation time (middle) on the first 500 problems in the APPS introductory dataset, using the APPS GPT-2 Transformer model.

	1@10	1@50	1@100
PG-TD	5.8	9.8	11.79
S+F	5.2	6	6.4
	5@10	5@50	5@100
PG-TD	6.8	11.6	12
S+F	5.6	6.4	7.6
	pass@10	pass@50	pass@100
PG-TD	7.19	13.8	14.79
S+F	7.39	10.8	12.8

Table 6: n@k and pass@k results of PG-TD and Sampling, evaluated on the first 500 APPS introductory problems.

	Pass Rate	Strict Accuracy
0 public test cases (the beam search baseline)	13.61	4.2
1 public test cases	21.24	7.6
3 public test cases	25.09	8.79
5 public test cases	26.59	10.8
Half of all test cases (10.44 public test cases on average)	32.11	12.8

Table 7: The performance of PG-TD using different number of public test cases, evaluated on the first 500 APPS introductory problems.

Using different numbers of public test cases. On the APPS dataset, we split the all the test cases evenly into public test cases and private test cases. To investigate how many test cases are enough for PG-TD to achieve a satisfactory performance, we consider using 1, 3, 5 public test cases and using the rest of the test cases as private test cases. The results are reported in Table 7, evaluated on the first 500 APPS introductory problems. We observe that even with a small number of public test cases, PG-TD is able to find programs with higher a pass rate and strict accuracy than beam search. This suggests that the Transformer can serve as a regularizer in code generation. With a small number of public test cases, PG-TD still generates programs similar to human-written programs without overfitting the pubic test cases.

Results on Codex. We have run both beam search and PG-TD algorithms using Codex (Chen et al., 2021a) as the backbone. We follow the "best practices" provided on the Codex guide to put the problem description into a block comment in the beginning, and ask Codex to complete the rest. The prompt looks like the following.

```
"""
Python 3
{problem description}
"""
```

As we expect, PG-TD helps Codex generate better codes. This further validates our claim that our method is model-agnostic and can help a pre-trained code generation Transformer generate better codes. Our results are reported in the Table 8. Due to the limits of the OpenAI APIs, we are only able to test the algorithms on a small set of data. We also show a concrete example where PG-TD outperforms beam search in Fig. 15.

Using sampling-based MCTS. We could use a sampling approach in the evaluation step of MCTS instead of using beam search, which would be a more faithful implementation of MCTS. We experimented with the two versions of PG-TD and find that using beam search indeed has a better performance (Table 9). The gap could be caused by that we only do sampling once to evaluate a new node. We could get a more accurate by sampling multiple times. However, that would make our algorithm more computationally expensive.

About failure modes. In Table 10, we report the average percentages of test cases where compilation errors and runtime errors occur. We use the error analysis utility provided in the original APPS codebase. As we expect, PG-TD executes the generated programs and finds the ones with higher pass rates, so it dramatically reduces the percentage of both types of errors.

About termination conditions. In Table 11, we find enabling early termination does (marginally) improve the algorithms' efficiency. We will include this implementation choice in the paper. We expect early termination would be significantly more helpful when the code generation algorithms can solve most of the problems (achieving pass rates of 1 on most of them).

Using PG-TD for code translation. A real-world use case of our algorithm is code translation. The goal of code translation is to translate a script from one programming language to another. In this problem, automatically generated test cases are used to determine if the translated codes are correct (Roziere et al., 2022). There exist free or commercial tools like EvoSuite, Jtest, etc. We conducted code translation experiments in the following table. We use TransCoder-ST (Roziere et al., 2022) as the backbone Transformer, and compare beam search with PG-TD. In PG-TD, it uses EvoSuite to generate public test cases, and use the pass rates on them as rewards. The translated codes are evaluated on a set of private test cases. We can see that in this problem, PG-TD also outperforms the beam search algorithm (Table 12).

Assets and licenses. The APPS dataset (Hendrycks et al., 2021) is released under MIT License [2]. /The CodeContests dataset (Li et al., 2022) is released under Apache License 2.0 [3]. The Huggingface Transformer (Wolf et al., 2019) is released under Apache License 2.0 [4].

B Planning for Other Code Generation Objectives

Besides the default reward function that optimizes the pass rate, we show the versatility of the proposed algorithm by setting two new objectives, code length penalty and comment encouragement.

Code length penalty. We make the planner to generate more concise codes by using the following reward function

$$\mathcal{R}_{\text{length}} = p + \lambda_1 \times e^{-l_c/t}, \tag{1}$$

where p is the average pass rate on the public test case set and l_c is the length of the code string. λ and t are hyperparameters that control the weight of code length penalty and are set to 0.1 and 20, respectively. As shown in the Figure 7, the generated solution becomes more concise and the code string length decreases from 187 to 78 while still passing all the test cases.

Comment encouragement. We can also generate solutions with more comments. We achieve this goal by using the following reward function

$$\mathcal{R}_{\text{comment}} = p + \lambda_1 \times e^{-l_c/t} + \lambda_2 \times \min(1, \frac{N_{cm}}{N_{max}}), \tag{2}$$

where N_{cm} is the number of "#" in the code string and λ_2 is set to 0.2, controlling the weights of the comment lines. If we simply add $\lambda_2 \times N_{cm}$ as the comment encouragement term, the planner would generate code with repeated meaningless comment lines to get more rewards. To handle this problem, we made the following two changes. First, we add the code length penalty term, $\lambda_1 \times e^{-l_c/t}$. Second, we upper-bound the rewards for the number of comment lines by setting the comment encouragement term to be $\lambda_2 \times \min(1, \frac{N_{cm}}{N_{max}})$, where $N_{max} = 5$ in the example shown in Figure 8. As shown in Figure 8, we can generate solutions with more comment lines with this designed reward function.

[2] https://github.com/hendrycks/apps/blob/main/LICENSE
[3] https://github.com/deepmind/code_contests/blob/main/LICENSE
[4] https://github.com/huggingface/transformers/blob/main/LICENSE

Model	Decoding Algorithm	Pass Rate	Strict Accuracy
APPS GPT-2 (1.5B)	Beam Search	9.38	0
APPS GPT-2 (1.5B)	PG-TD	30.99	2.38
Codex (code-davinci-002)	Beam Search	34.56	9.52
Codex (code-davinci-002)	PG-TD	47.17	23.80

Table 8: Comparison between using different Transformers as backbone for code generation, evaluated on the first 42 APPS introductory problems.

Decoding Algorithm	Pass Rate	Strict Accuracy
PG-TD using Beam Search	32.757	13.2
PG-TD using Sampling	15.69	5

Table 9: PG-TD using beam search vs. sampling in the evaluation step, evaluated on the first 500 APPS introductory problems.

	Compilation Error (%)	Runtime Error (%)
APPS GPT-2 (1.5B), Beam Search	5.58	32.95
APPS GPT-2 (1.5B), Sampling + Filtering	4.68	27.38
APPS GPT-2 (1.5B), PG-TD	1.93	19.5

Table 10: Failure case analysis. The percentages are averaged over the APPS introductory dataset.

	Computation Time (sec.)
PG-TD without early termination (as in paper)	1177.99
PG-TD with early termination	986.37
Sampling without early termination (as in paper)	2154.23
Sampling with early termination	2081.25

Table 11: Computation times with and without early stopping, using APPS GPT-2 (1.5B).

	Python $\rightarrow$ Java	CPP $\rightarrow$ Java
Beam Search	58.2	68.0
PG-TD	63.5	70.6

Table 12: Pass rates of applying our method for code translation, where test cases are automatically generated. The baseline results are copied from the original paper.

Problem Statement
Takahashi loves palindromes. Non-palindromic strings are unacceptable to him. Each time he hugs a string, he can change one of its characters to any character of his choice. Given is a string S. Find the minimum number of hugs needed to make S palindromic.
Constraints
(1). S is a string consisting of lowercase English letters.
(2). The length of S is between 1 and 100 (inclusive).
Input
Input is given from Standard Input in the following format:S.
Output
Print the minimum number of hugs needed to make S palindromic.
Sample Test Input 1
redcoder
Sample Test Output 1
1
For example, we can change the fourth character to o and get a palindrome redooder.

```
# cook up a palindrome
s=input()
n=len(s)
c=0
if n%2==0:
        for i in range(n//2):
                if s[i]!=s[n-i-1]:
                        c+=1
        print(c)
else:
        for i in range(n//2):
                if s[i]!=s[n-i-1]:
                        c+=1
        print(c)
```

Default reward function (Pass Rate: 1.00, code string length: **187**).

```
s=input()
ans=0
for i in range(len(s)//2):
        if s[i]!=s[~i]: ans+=1
print(ans)

```

Reward function with code length penalty (Pass Rate: 1.00, code string length: **78**).

Figure 7: A code generation example with code length penalty. The problem description is shown on the top. The programs generated by default PG-TD and PG-TD with code length penalty are shown at the bottom. We can generate more concise code by code length penalty reward.

Problem Statement
$A + B$ is often used as an example of the easiest problem possible to show some contest platform. However, some scientists have observed that sometimes this problem is not so easy to get accepted. Want to try?
Input
The input contains two integers a and b $(0 \leq a, b \leq 10^3)$, separated by a single space.
Output
Output the sum of the given integers.
Sample Test Input 1
5 14
Sample Test Output 1
19
Sample Test Input 2
381 492
Sample Test Output 2
873

```
a, b = map(int, input().split())
print(a + b)

```

Default reward function.

```
#!/usrbin/env python3
# coding =utf-8 #py3 runs on py2
a,b=map(int,input().split())
print(a+b)
```

Reward function with comment encouragement (code with comments).

Figure 8: A code generation example with code comment encouragement. We can generate codes with more comments by using the code comment encouragement reward.

C ALGORITHM DETAILS

C.1 PG-TD

In this section, we describe our PG-TD algorithm with more details, following the Monte-Carlo Tree Search (MCTS) steps. We show our framework in Fig. 10 again for convenience. The algorithm maintains a search tree structure where nodes correspond to states and edges correspond to actions. For each node s, it maintains the number of times that it has been visited, denoted by $s.visits$. For a state, action pair, s, a (corresponding to an edge in the tree), the algorithm maintains a function $Q(s, a)$, which is the maximum value obtained by starting in s and taking action a. Conventionally, $Q(s, a)$ is the *average* return. Since our the code generation problem is deterministic (that is, there are no stochastic transitions), we find the performance is better if we keep track of the best program generated from a node.

First, the algorithm **selects** a node for expansion. It starts from the root node, selects substrees recursively, until it reaches a node which is not fully extended (whose children are fewer than k_{max}). The common node selection criterion is upper confidence bound (UCB) (Kocsis & Szepesvári, 2006), which leverages between exploiting known-to-be-good states and exploring less-visited states,

$$\text{UCB}(s, a) = Q(s, a) + \beta\sqrt{\frac{\log(s.visits)}{s'.visits}}, \tag{3}$$

where s' is the state reached by taking action a in s; β is a hyperparameter. We set $\beta = 1$ in our experiments. The UCB_SELECT function in Alg. 1 returns the next state that has the highest UCB value,

$$\text{UCB\_SELECT}(s) = \arg\max_a \text{UCB}(s, a). \tag{4}$$

Then the algorithm **expands** a node using GENERATE_SEQ and **evaluates** the new node using GENERATE_SEQ, as we described in the main paper. Finally, the estimated value is **backpropagated** to its parents recursively until it reaches the root and update their Q values. Suppose the estimated value is v. Then for all state, action pairs, s', a', along the trajectory in the tree to reach the new state, $Q(s', a')$ is updated accordingly using the new value: $Q(s', a') \leftarrow \max(Q(s', a'), v)$.

Comparison with the standard MCTS algorithm. It is worth noting that our implementation of MCTS in PG-TD is different from the standard MCTS algorithm. If we use the standard MCTS algorithm, it is computationally intractable to search the large space of possible programs to find high-quality ones.

For comparison, we visualize the standard MCTS algorithm in Fig. 9. In the expansion step, MCTS selects the next available action in the action set, and adds the state that can be reached by following the action. In the example, action "`b`" is taken, and the new state that is added to the tree is "`ab`". In the evaluation step, MCTS evaluates the new state by executing a random policy from the new state and computes the value of the policy.

It is impractical to apply the standard MCTS algorithm to domains with large state spaces or large action spaces, as we cannot afford to try all possible actions in the expansion step. A random policy in the evaluation step also has a high variance in estimating the value of the new state. In PG-TD, we use a pre-trained Transformer to resolve these limitations of the standard MCTS algorithm.

Transformer generation functions. Our framework has a Transformer model that is pretrained on competitive coding datasets, and provides the GENERATE_SEQ function that generates most-likely sequences from a partial program, and TOP_K that generates the most-likely next tokens given a partial program. We describe the implementations of these two functions in Alg. 4 and Alg. 5. They both make use of $P_{\text{Transformer}}(a|s)$, the probability that token a is the next token given the partial program s, which is readily provided in a pre-trained Transformer model. Specifically, TOP_K selects the top-k most-likely tokens according to $P_{\text{Transformer}}$. GENERATE_SEQ follows the standard Transformer beam search algorithm to generate complete programs.

C.2 EFFICIENT IMPLEMENTATION BY INFORMATION SHARING

As discussed in the main paper, there may be calls to the Transformer generation functions the perform redundant computations. We have described tree structure caching in the main paper. We provide more details on *sequence caching* in this section.

Sequence caching. The goal of sequence caching is to reduce the computational cost of the evaluation step of PG-TD. In the evaluation step, the Transformer beam search algorithm is called to generate a complete program from the current node. These complete programs and their rewards are both cached and used in a future iteration.

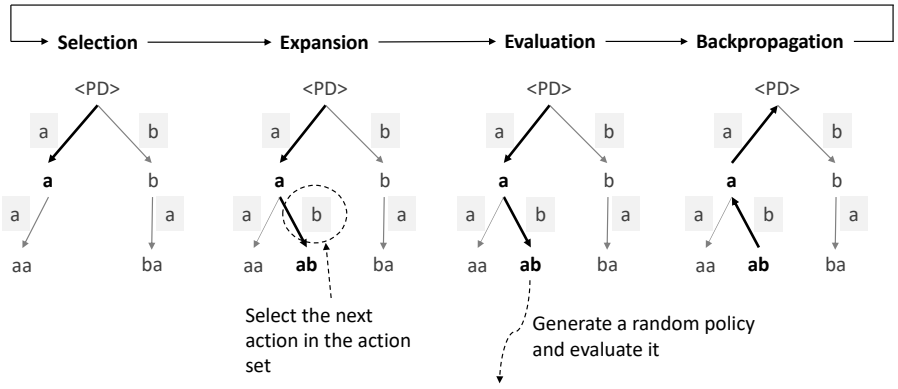


Figure 9: The standard Monte-Carlo tree search algorithm, without using a pre-trained Transformer. <PD> stands for problem description.

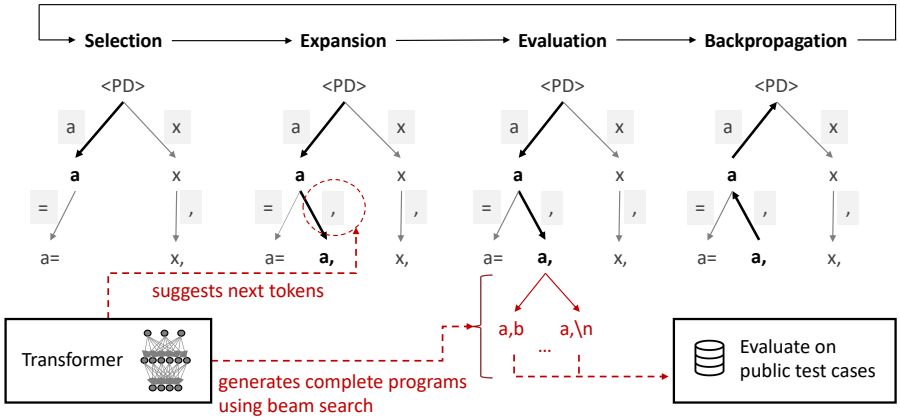


Figure 10: Our proposed framework, PG-TD, repeated here for convenience.

However, we can only use sequence caching for beam size $b = 1$. When $b > 1$, using cached sequences in the previous iterations may be suboptimal. Let's consider the example in Figure 12, where $b = 2$. Starting from "`a,`" (the left figure), suppose the beam search algorithm returns a sequence by searching two beams with the prefixes of "`a,b`" and "`a,c`", respectively (the bold lines). In the next iteration (the right figure), to evaluate "`a,b`" using the beam search algorithm, we may not use the sequence found in the previous iteration, which *only searches the subtree of "`a,b`" with one beam*. In other words, we may find a sequence with a higher likelihood by re-running the beam search algorithm with $b = 2$ starting from "`a,b`".

In Alg. 4 and Alg. 5, we described how GENERATE_SEQ and TOP_K take advantage of the caching mechanism, highlighted in blue colors. In Alg. 6, we described how we clean the cache so that the memory it takes does not increase without bound as we execute PG-TD. Alg. 6 is called at the beginning of each function call of PG-TD.

We also cache the search tree between different function calls of PG-TD, which is the same design choice used in Silver et al. (2017). The tree structure built in one iteration will be kept and used for following iterations. So the following iterations can search even deeper in the tree.

C.3 Baseline Algorithms

Sampling + Filtering. We have described the sampling algorithm in the main paper. The pseudocode is shown in Alg. 7.

Sequential Monte-Carlo-Guided Transformer Decoding. Sequential Monte-Carlo-Guided Transformer Decoding (SMCG-TD) is another algorithm that we compare PG-TD with, which also uses Transformer to evaluate partial programs. The pseudocode is shown in Alg. 8. One can think of it from a view of genetic algorithms. SMCG-TD maintains a group of partial programs, called a *generation*, and updates them iteratively. For each partial program in the generation, we sample a next token using the Transformer and append it to the

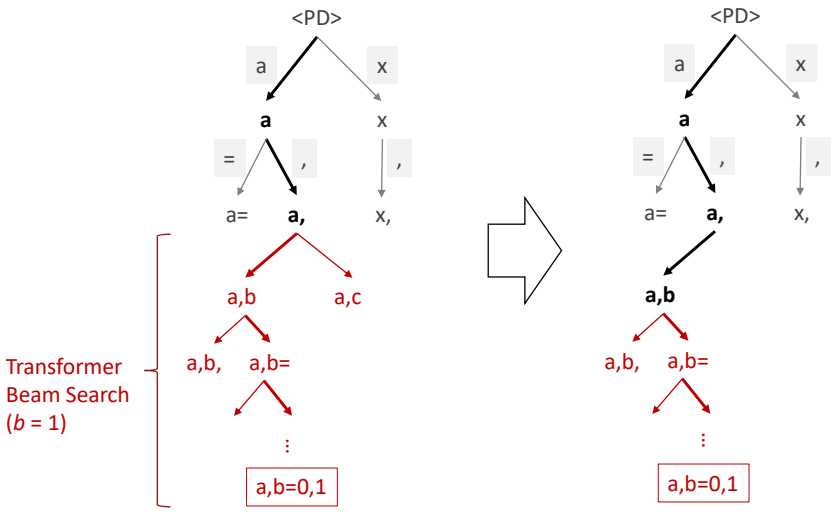


Figure 11: Illustration of sequence caching. a most-likely sequence generated in an iteration of PG-TD (left) may be used for a future iteration (right).

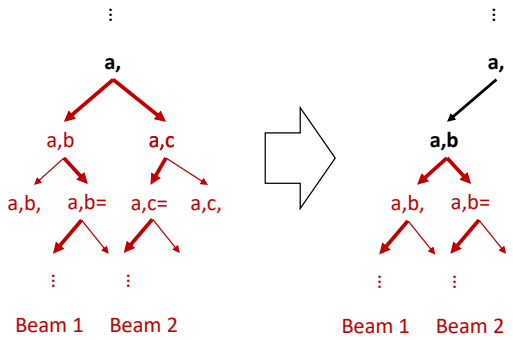


Figure 12: Illustration of an example where sequence caching may not return the correct sequence when $b > 1$.

partial program, and put it in a new generation. We then evaluate each partial program in the new generation in the same way as PG-TD: We use the Transformer to generate a complete program from the partial program and compute its reward (pass rate). We then use these rewards as the partial programs' fitness scores. At the end of an iteration, we re-sample from the partial programs according to their fitness. So partial programs with higher fitness scores are more likely to remain in the generation. In the end, the algorithm outputs the complete program it finds that has the largest reward.

One may notice that SMCG-TD resembles PG-TD on a higher level. It also uses the Transformer and the public test cases to decide which next token to generate in the code generation process. The key difference is that SMCG-TD uses a sampling-based approach and PG-TD uses a tree search algorithm. We are able to increase the number of rollouts in PG-TD to make it do multi-step look-ahead search, which is what SMCG-TD incapable of. So in our experiments, the performance of SMCG-TD only approximates the default setting of PG-TD ($k_{max} = 3, rollouts = 3$) where we do one-step look-ahead search in the tree search algorithm.

It is worth noting that sequence caching is also implemented for SMCG-TD when it calls GENERATE_SEQ. So even if multiple partial programs in a generation can be the same, the algorithm would not call the Transformer to generate a complete program from the same prefix more than once, which effectively reduces the number of times that it needs to call the Transformer and saves the computation time.

Algorithm 3 The PG-TD algorithm. Function calls to the pre-trained Transformer are highlighted in red. Caching-related codes are in blue.

Require: $root$: the current state; $rollouts$: the number of MCTS rollouts; k_{max}: the maximum number of children of any node; b: the number of beams for Transformer beam search

UPDATE_CACHE($root$) ▷ Alg. 6
for $i \leftarrow 1, 2, \ldots, rollouts$ **do**
  $node \leftarrow root$
  # Selection
  while $|node.children| = k_{max}$ **do**
    $node \leftarrow$ UCB_SELECT($node.children$)
  end while
  # Expansion
  $children\_num \leftarrow |node.children|$
  $act \leftarrow$ TOP_K($node, k = children\_num + 1$) ▷ Alg. 4
  $new\_node \leftarrow$ the state reached by taking action act in $node$
  Add new_node to $node.children$
  # Evaluation
  $program\_list \leftarrow$ GENERATE_SEQ(new_node, b) ▷ Alg. 5
  $r \leftarrow \max_{p \in program\_list}$ GET_REWARD(p)
  # Backpropagation
  Update and the values of new_node and its ancestors in the tree with r
end for
return the action with the highest value starting from $root$

Algorithm 4 TOP_K. Caching-related codes are in blue.

Require: s: starting partial program; k: returning the k-the most likely token.

if $s \in top\_k\_cache.keys()$ **then**
  $act \leftarrow$ the k-st most-likely token from $top\_k\_cache[s]$
end if
$sorted\_tokens \leftarrow$ sort all tokens by $P_{\text{Transformer}}(\cdot|s)$ in descending order
$top\_k\_cache[s] \leftarrow sorted\_tokens$
return $sorted\_tokens[k]$

C.4 FINETUNING

In our experiments, we use the samples generated by the PG-TD to further finetune the Transformer models. We then generate solutions using the finetuned Transformer model by running beam search, and observe improvement on both pass rate and strict accuracy compared with running beam search using the original model.

One challenge in our design is that solutions that PG-TD generates have different pass rates. We could use only solutions with pass rate of 100%. However, that largely limits the number of samples we can use for finetuning. We instead collect all solutions that have pass rates larger than 80%. To make the Transformer aware of the different pass rates of training samples, we use the loss function similar to Le et al. (2022),

$$\mathcal{L}(\theta) = -\mathbb{E}_{W \sim p_\theta}[r(W) \nabla_\theta \log p_\theta(W|PD)], \tag{5}$$

where W denotes a complete program; $r(W)$ is the pass rate of the program; PD is the program description. In future work, we will investigate different loss functions and compare their effectiveness in improving the Transformer model's performance.

Algorithm 5 GENERATE_SEQ. The Transformer beam search algorithm. Caching-related codes are in blue.

Require: s: starting partial program; b: beam size; k: the top-k next tokens to consider.

if $b = 1$ and there exists $cached\_state \in sequence\_cache.keys()$ (s matches the prefix of $cached\_state$) **then**
  return $sequence\_cache[cached\_state]$
end if
$candidates \leftarrow [s]$
$t \leftarrow 0$
$outputs \leftarrow$ NEW_LIST()
while $|candidates| > 0$ and $t < max\_steps$ **do**
  $new\_candidates \leftarrow$ NEW_LIST()
  for $s \in candidates$ **do**
    for $k' \leftarrow 1, \ldots, k$ **do**
      $a \leftarrow$ TOP_K(s, k')
      $s' \leftarrow$ CONCATENATE(s, a)
      add s' to $new\_candidates$
    end for
  end for
  compute the likelihood of all the partial programs in $new\_candidates$, and only keep the top-b partial programs in $new\_candidates$ and drop the rest
  move complete programs, if any, from $new\_candidates$ to $outputs$
  $t \leftarrow t + 1$
end while
if $b = 1$ **then**
  $sequence\_cache[cached\_state] \leftarrow outputs$
end if
return $outputs$

Algorithm 6 UPDATE_CACHE.

Require: s: the current partial program.

for $cached\_state \in sequence\_cache.keys()$ **do**
  if s does not match the prefix of $cached\_state$ **then**
    delete $sequence\_cache[cached\_state]$
  end if
end for
for $cached\_state \in top\_k\_cache.keys()$ **do**
  if s does not match the prefix of $cached\_state$ **then**
    delete $top\_k\_cache[cached\_state]$
  end if
end for

Algorithm 7 Sampling + Filtering.

Require: $sample\_num$: the number of samples to generate using Transformer.

for $i \leftarrow 1, 2, \ldots, sample\_nums$ **do**
  $p_i \leftarrow$ TRANSFORMER_SAMPLE$(\langle PD \rangle, k = 5)$
  $rewards[i] \leftarrow$ GET_REWARD(p_i)
end for
return p_i with the largest $rewards[i]$

Algorithm 8 Sequential Monte Carlo-Guided Transformer Decoding (SMCG-TD).

Require: pop_size: the population size
$generation \leftarrow [\langle PD \rangle] * pop\_size$ # initialize the population with the problem description
$fitness \leftarrow [1.0/pop\_size] * pop\_size$ # uniform distribution
$t \leftarrow 0$
$complete\_programs \leftarrow$ NEW_LIST()
while $|population| > 0$ and $t < max\_steps$ **do**
    $new\_generation \leftarrow$ NEW_LIST()
    $new\_fitness \leftarrow$ NEW_LIST()
    for $i \leftarrow 1, \ldots,$ LEN($generation$) **do**
        $s \leftarrow$ sampled from $generation$, where $generation[j]$ is selected with the probability of $fitness[j]$
        $a \leftarrow$ sampled from $P_{Transformer}(\cdot|s)$
        $s' \leftarrow$ CONCATENATE(s, a)
        if $s'[-1] =$ `<|endoftext|>` **then**
            # this sample finishes generation, save to output
            add s' to $complete\_programs$
        else
            add s' to $new\_generation$
            $p \leftarrow$ GENERATE_SEQ$(s', b = 1)$
            $r \leftarrow$ GET_REWARD(p)
            add r to $new\_fitness$
        end if
    end for
    normalize $new\_fitness$
    $generation \leftarrow new\_generation$
    $fitness \leftarrow new\_fitness$
    $t \leftarrow t + 1$
end while
return the program in $complete\_programs$ that has the largest reward

D ILLUSTRATIVE EXAMPLES

D.1 CODE GENERATION EXAMPLES

In this section, we provide more qualitative examples generated by our PG-TD and the baseline methods, shown in Figure 13 and Figure 14. Our PG-TD algorithm generates solutions that have higher pass rates than the baseline methods.

Problem Statement

Takahashi is meeting up with Aoki. They have planned to meet at a place that is D meters away from Takahashi's house in T minutes from now. Takahashi will leave his house now and go straight to the place at a speed of S meters per minute. Will he arrive in time?

Constraints

(1). $1 \leq D \leq 10000$
(2). $1 \leq T \leq 10000$
(3). $1 \leq S \leq 10000$
(4). All values in input are integers.

Input

Input is given from Standard Input in the following format: $D\ T\ S$

Output

If Takahashi will reach the place in time, print Yes*; otherwise, print* No*.*

Sample Test Input

1000 15 80

Sample Test Output

Yes

It takes 12.5 minutes to go 1000 meters to the place at a speed of 80 meters per minute. They have planned to meet in 15 minutes so he will arrive in time.

```
# cook your dish here
try:
  a=list(map(int,
  ↪  input().split(
  ↪  )))
  if(a[2]==1):
    print("Yes")
  else:
    print("No")
except:
  pass
```

Beam Search (Pass Rate: 0.125).

```
# cook your dish here
d,t,s=list(map(int,
↪  input().split
↪  ()))
if(d==t):
  if(s*(s+1)/s==t):
    print("Yes")
  else:
    print("No")
else:
  if(s*(s+1)/s>t):
    print("Yes")
  else:
    print("No")
```

Sampling + Filtering (Pass Rate: 0.625).

```
n, t, s=map(int,
↪  input().split())
t1=n / s
if t1 <= t:
  print("Yes")
else:
  print("No")
```

PG-TD (Pass Rate: 1.00).

Figure 13: A code generation example for competitive programming.The problem description is shown on the top. The programs generated by baseline algorithms and our PG-TD algorithm are shown at the bottom.

Problem Statement

Given is an integer r. How many times is the area of a circle of radius r larger than the area of a circle of radius 1? It can be proved that the answer is always an integer under the constraints given.

Constraints

(1). $1 \leq r \leq 100$.

(2). All values in input are integers.

Input

Input is given from Standard Input in the following format: r.

Output

Print the area of a circle of radius r, divided by the area of a circle of radius 1, as an integer..

Sample Test Input

2

Sample Test Output

4

The area of a circle of radius 2 is 4 times larger than the area of a circle of radius 1. Note that output must be an integer - for example, 4.0 will not be accepted.

```
# cook your dish here
r=int(input())
if(r>=2*r):
        print(r*r*2)
else:
        print(0)
```

Beam Search (Pass Rate: 0.00).

```
# cook your dish here
r=int(input())
if(r>=2*r):
        print(r*r*2)
else:
        print(0)
```

Sampling + Filtering (Pass Rate: 0.00).

```
# cook your dish here
r=int( input() )
print(2*(r**2)-r**2)
```

PG-TD (Pass Rate: 1.00).

Figure 14: A code generation example for competitive programming. The problem description is shown on the top. The programs generated by baseline algorithms and our PG-TD algorithm are shown at the bottom.

Problem Statement
You are given an array a consisting of n integer numbers.
Let instability of the array be the following value: $\max_{i=1}^{n} a_i - \min_{i=1}^{n} a_i$.
You have to remove exactly one element from this array to minimize instability of the resulting $(n-1)$-elements array. Your task is to calculate the minimum possible instability.
Input
The first line of the input contains one integer n ($2 \le n \le 10^5$) — the number of elements in the array a.
The second line of the input contains n integers $a_1, a_2, \ldots, a_n$ ($1 \le a_i \le 10^5$) — elements of the array a.
Output
Print one integer — the minimum possible instability of the array if you have to remove exactly one element from the array a.
Examples
Input
4
1 3 3 7
Output
2
Input
2
1 100000
Output
0
Note
In the first example you can remove 7 then instability of the remaining array will be $3 - 1 = 2$.
In the second example you can remove either 1 or 100000 then instability of the remaining array will be $100000 - 100000 = 0$ and $1 - 1 = 0$ correspondingly.

```
n = int(input())
a = list(map(int, input().split()))

print(max(a) - min(a) - 1)
```

Beam search (Pass Rate 0.00).

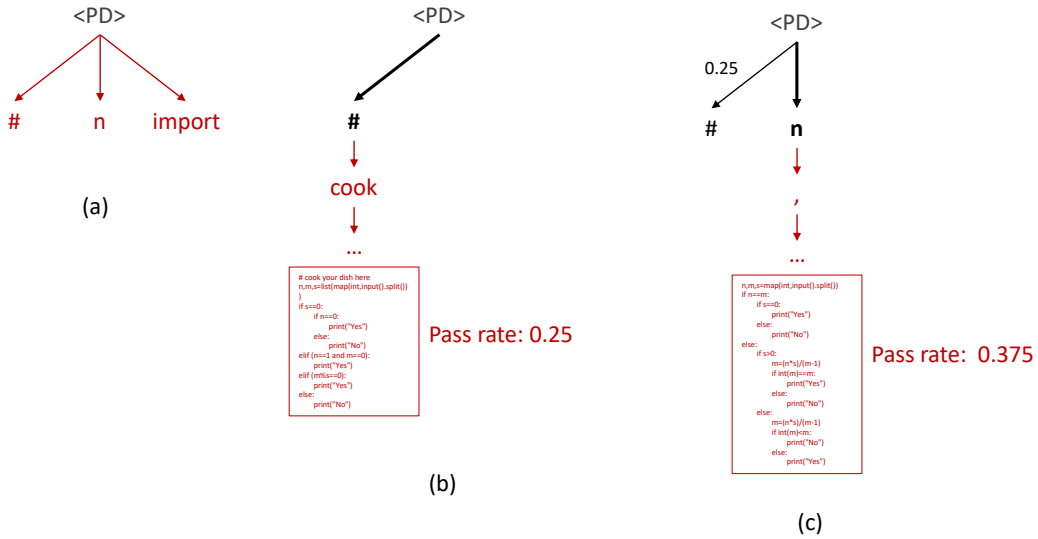
```
def main():
    n = int(input())
    a = list(map(int, input().split()))
    a.sort()
    print(min(a[n-1] - a[1], a[n-2] - a[0]))

if __name__ == "__main__":
    main()
```

PG-TD (Pass Rate 1.00).

Figure 15: A code generation example using Codex (code-davinci-002). Beam search finds an incorrect solution while PG-TD finds a correct solution. This result further confirms that PG-TD is model-agnostic and could be effective with different backbone Transformers (GPT-2 (1.5B), GPT-Neo (2.7B) and Codex).

D.2 ILLUSTRATIONS OF STEP-BY-STEP EXECUTION OF PG-TD

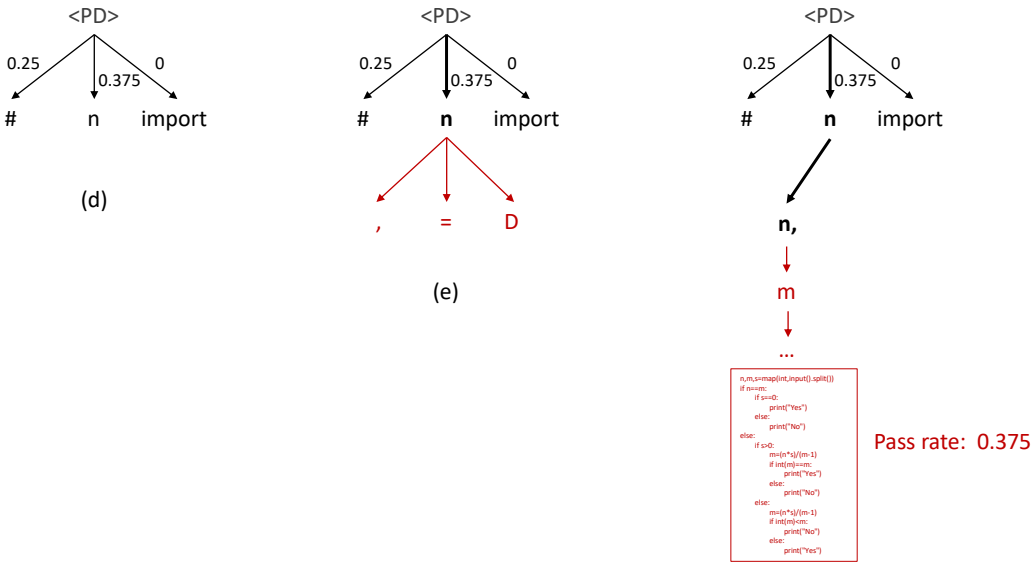
In the example shown in Fig. 13, we see that PG-TD outputs the first tokens ("`n, t, ...`") differently from the beam search ("`# cook your ...`"). It means that after evaluating the pass rates of the potentially generated programs, our algorithm disagrees with the Transformer beam search algorithm and outputs different tokens. We use this problem as an example to visualize the execution of PG-TD. We consider the setting where the maximum children of any node (k_{max}) is 3, the number of rollouts is 6, and the number of beam size (b) is 1.



(a) We start with decoding from the problem description, `<PD>`. Initially, there is only the root node, which is the problem description. To expand this node, the algorithm calls TOP_K to ask the Transformer to suggest the most-likely tokens. In this case, Transformer suggests `'#', 'n', 'import'`, in descending order of their likelihood.

(b) The algorithm adds the most-likely token, `'#'`, to the tree. To evaluate this node, the algorithm calls GENERATE_SEQ to generate a complete program, and run it on test cases. The pass rate of the generated program is 0.25 (it passes 25% of the public test cases).

(c) In the next iteration, the algorithms adds the second most-likely next token, `'n'`, to the tree. Similarly, the algorithm generates a complete program from this new node and runs it on the test cases. Its pass rate is 0.375.

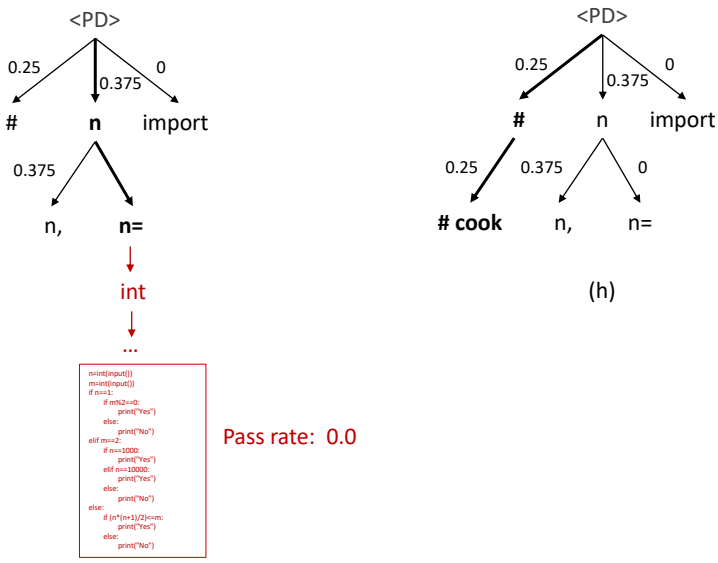


(d) The process is similar for the third most-likely token, `'import'`. The program generated starting from it has 0 pass rate.

After the first three iterations, three nodes are added to the tree and their values are computed.

(e) Now the algorithm selects `'n'` for expansion, as it has the highest value. The algorithm calls TOP_K to get the most-likely tokens. In this case, they are `','`, `'='`, `'D'`.

(f) The algorithm adds `'n,'` to the tree. The program generated from it has the pass rate of 0.375.



(g) The algorithm selects `'n'` again and adds `'n='` to the tree. The program generated from it has the pass rate of 0.

(h) The algorithm selects `'#'` for expansion and adds `'# cook'` to the tree. It no longer selects `'n'` for expansion in this iteration as it believes the subtree of `'n'` is well explored, and it instead explores the substree of `'#'`.

Now the algorithm has finished six iterations of rollouts. One can verify that six nodes have been added to the tree. After the tree search, the algorithm will return the token `'n'` as the action it believes to be optimal. Our algorithm disagrees with the most-likely token suggested by the Transformer, as it finds that it can generate a program with a higher pass rate by outputting a different token (the second most-likely token suggested by the Transformer).

E MORE DISCUSSIONS

Potential benefits. We have shown that our framework can not only help achieve higher pass rates compared with using only the Transformer generation process, but also generate codes that optimize different objectives. These are done without fine-tuning the pre-trained Transformer models. Our framework makes a pre-trained Transformer more versatile and adapts it to different tasks by designing different reward functions used by the planning algorithm. If we re-train or fine-tune a Transformer like GPT-2 used in this paper, we will need to train billions of parameters (Radford et al., 2019). Our approach potentially saves the computational cost and energy consumption that is needed to train or fine-tune Transformers for different tasks.

Potential negative social impacts. Automatic code generation makes it easier for anyone to generate programs that meets a specification. Our hope is that this development will relieve the burden of software engineers and enable AI-assisted programming or even end-to-end automatic programming. However, it may also make it easier to develop malwares. Anyone that can specify a malicious goal in a natural language can use an automatic code generation tool to generate codes that meet the goal. When automatic code generation techniques become more mature, we may need a separate module that screens the natural language description and rejects the code generation requests that can lead to harmful codes.