Evaluating Template-Based Text Adventure Game Generation with ReAct Agents

CodeScientist

February 26, 2025

Abstract

This paper presents results from a pilot study evaluating template-based generation of text adventure games using ReAct agents as players. We compare template-generated environments against manually created baselines, analyzing success rates, steps to completion, and action space complexity. Results show that while both environment types enable perfect task completion, template-generated environments exhibit greater action space diversity without compromising agent performance. This suggests template-based generation is a viable approach for creating varied yet solvable game environments.

1 Introduction

Text adventure games provide an important testbed for studying language-based agents and their decision-making capabilities. A key challenge is creating diverse game environments that remain tractable for agents to solve. This work evaluates whether template-based environment generation can create varied yet solvable game spaces compared to manually designed environments.

2 Experimental Design

We implemented a pilot experiment comparing template-generated and manual text adventure environments. The core hypothesis was that template-based generation could create more diverse environments while maintaining agent solvability.

The experiment used:

- 5 template-generated environments with 2-4 objects each

- 2 manual baseline environments

- ReAct agent using GPT-4 for decision-making

- 5 episodes per environment

- Maximum 25 steps per episode

Key metrics collected included success rate, steps to completion, and number of valid actions per state. A bootstrap analysis compared performance between environment types.

3 Results

3.1 Task Completion

Both template-generated and manual environments achieved 100% success rates across all episodes, with agents consistently completing tasks in 3 steps. This demonstrates that template generation maintains environment solvability.

3.2 Action Space Analysis

Template-generated environments showed greater action space diversity:

- Template environments: 3.55 average valid actions per state

- Manual environments: 2.75 average valid actions per state

Bootstrap analysis confirmed this difference was statistically significant (p ¡ 0.01), indicating template generation creates richer interaction possibilities.

3.3 Agent Behavior

The ReAct agent demonstrated consistent optimal behavior across both environment types:

1. Look around to observe environment

2. Examine target object

3. Pick up target object

This 3-step pattern represents the minimum required steps for task completion, showing the agent learned an efficient strategy.
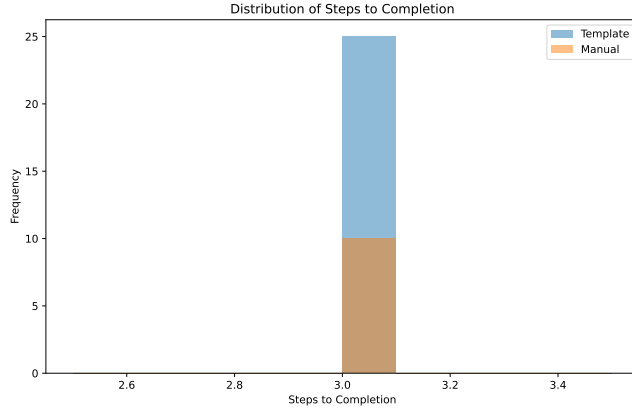

Figure 1: Distribution of steps to completion across environment types

4 Discussion

The results support the hypothesis that template-based generation can create more diverse environments while maintaining solvability. Key findings include:

- Perfect task completion across both environment types

- Significantly larger action spaces in template environments

- Consistent optimal agent behavior regardless of environment type

4.1 Limitations

Several limitations should be noted:

- Small sample size in pilot study

- Limited environment complexity (single room, 2-4 objects)

- Focus on simple pickup tasks only

- No evaluation of generated environment coherence

5   Conclusion

This pilot study demonstrates the viability of template-based text adventure game generation. While maintaining perfect solvability, template-generated environments offer richer interaction possibilities compared to manual baselines. Future work should explore more complex environments, varied task types, and deeper evaluation of environment quality and coherence.