Cybernetic Reinforcement Learning Agent for the Wumpus World Environment

Cesar Pulido[1], Cristian Romero[2]

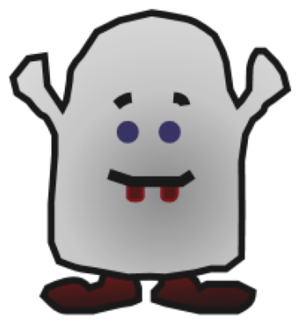
Universidad Distrital Francisco José de Caldas[1,2]

UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS

UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS

Introduction

The wumpus world is a classical problem in Artificial Intelligence field of study for its simplicity and, at the same time, its complications if viewed from the wrong approach. The environment of this world consists of a grid world with many dangers, such as wells and creatures called Wumpus, as well as gold, which is the target object that the agent must find, obtain and bring it to the initial position. Agent receives local information and must interpret this information in order to survive and achieve the goal. The key challenge is to make learning as adaptive, efficient and intelligent as possible in agent behavior through iterative improvements in perception and control. It has to be robust and able to operate in dynamic and unpredictable environments, while avoiding undesirable behaviors such as infinite loops or unsafe exploration.



Goal

We intend to implement an adaptive agent capable of navigating and making intelligent decisions using principles from system dynamics and reinforcement learning. The goal of this project is to combine dynamical systems analysis, feedback loops and Deep-Q-Learning algorithms to develop an all-around agent for this classical problem.

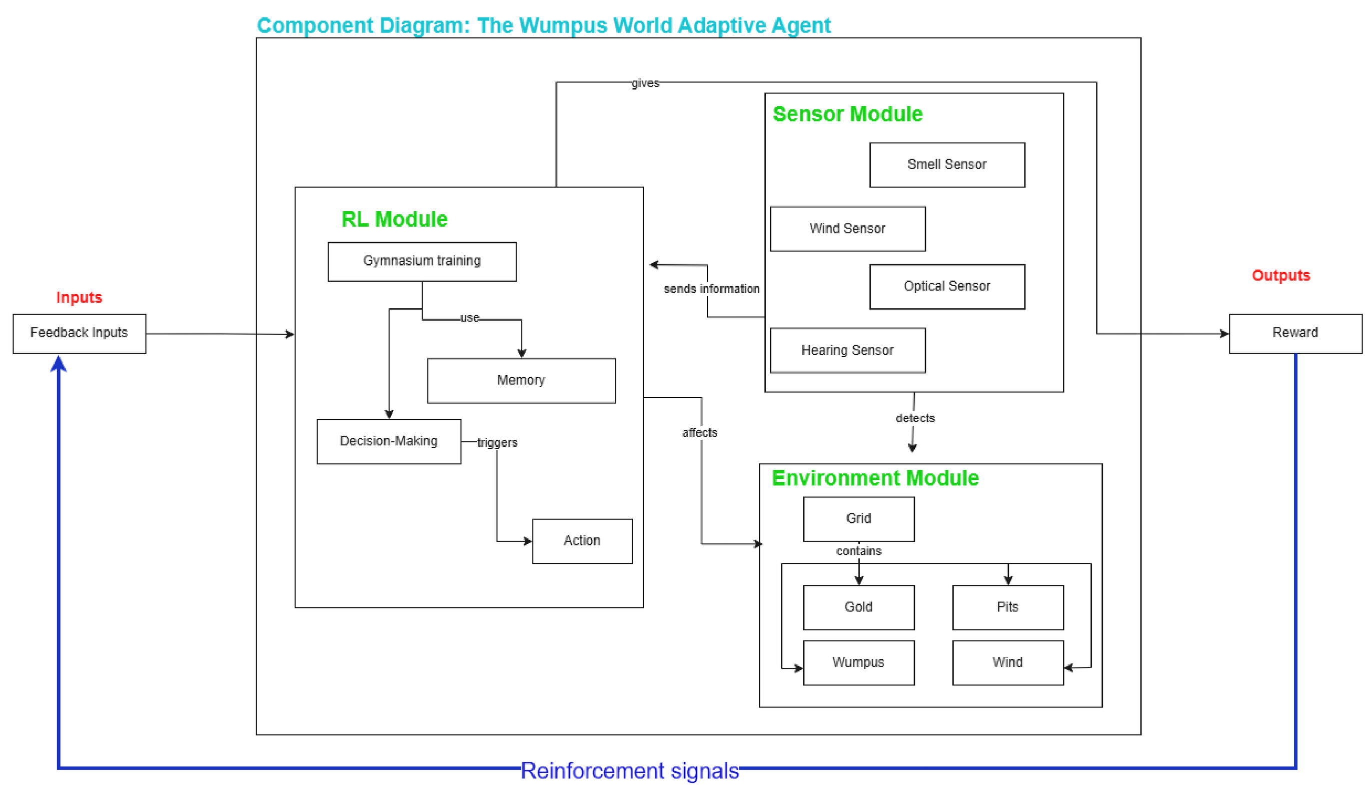
Experiments

We propose a series of simulation-based experiments to evaluate the expected learning and decision-making performance of the agent under different conditions. Each test scenario is designed to explore how the agent would behave across increasing levels of uncertainty and environmental complexity.

Test ID	Grid Type	Hazards Config.	Reward Model
T1	Static (4x4)	2 pits, 1 Wumpus	Sparse + terminal
T2	Randomized hazards	3 pits, 1 Wumpus	Shaped + penalty
T3	Variable start state	3 pits, 1 Wumpus	Shaped + penalty

The experiments are planned to simulate training over multiple episodes using a Deep Q-Network (DQN) with e-greedy exploration, experience replay, and a target network. Unit tests will be implemented to validate the functional correctness of environment initialization, percept logic, and reward handling. Integration and acceptance tests will verify whether the agent can complete its objective—retrieving the gold and returning to the starting position—under various hazard layouts. The goal of these experiments is to analyze learning progression using key performance indicators such as cumulative reward, episode length, and overall success rate. These metrics will help estimate the convergence and stability of the policy once implemented.

Proposed solution



Our approach is to use RL algorithms for adaptive behavior, define custom observation and action spaces, implement environment logic like event-based rewards, and integrate seamlessly with learning algorithms such as Deep Q-Networks. As the diagram shows, the general architecture is based on three main modules, which at each time step result in a reward that will be useful for agent learning.

Results

Since the implementation is still in development, we propose a set of expected behaviors and measurable outcomes based on theory and system design.

Conclusions

Wumpus provides a valuable environment for studying the interaction between system dynamics and reinforcement learning. The proposed system aims to model intelligent behavior in constrained and uncertain environments by combining reinforcement learning and network principles such as feedback, control and adaptation. The architecture includes modular components for sensing, learning and decision-making, with a focus on feedback loops and robustness.

References

[1] M. Genesereth and N. Nilsson, *Logical Foundations of Artificial Intelligence*. Morgan Kaufmann, 1987.

[2] V. Mnih, K. Kavukcuoglu, D. Silver, A. Rusu, et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[3] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, pp. 279–292, 1992.