**Introduction**

A common task in multi-agent systems is guessing an agent’s ‘type’ based on observations of their actions. Imagine a team of doctors diagnosing a patient – the doctors do not know if the patient has a common cold (Type 1) or ebola (Type 2). Based on the characteristics of each disease and observations of the patient (such as diagnostic reports), the doctors can guess the patient’s type, and can start the correct treatment plan. Knowing the patient’s type allows the doctors to make advantageous decisions and maximize utility.

However, complications can arise as the doctors attempt to diagnose the patient. Doctors may have learned slightly different knowledge about each disease in the past, leading doctors to have different beliefs about the true characteristics of each disease. Doctors may also receive different observations, with the radiologist interpreting the radiology report, the cardiologist focusing on the cardiology report, etc. In diagnosing the patient, the doctors must combine their individual estimates of the probability of each disease.

In this paper, we discuss a Bayesian approach to estimating types, as well as strategies for combining type estimates from multiple observers. We use a simple grid domain where each agent type has a distinct goal position on the board, and measure the accuracy of an observer guessing this agent type by observing movements on the board. We begin with the case of a single observer, and measure the accuracy and speed of type identification across goal and board configurations. We then extend to the case of multiple observers, studying the accuracy of type identification with different numbers and strategies of observers.

**Related Work**

**Design and Approach**

This paper uses a simple grid domain inspired by Colored Trails. The playing agent is placed at some location on the grid, and knows the true position of the goal on the grid. The goal of the playing agent is to reach the goal as quickly as possible, which it does by solving an MDP for the optimal path. The observing agents are not on the grid, but they can observe the playing agent’s movements. Each observing agent has a prior belief about the goal position for each agent type.

*Moving Around the Board*

The choice of a simple 2-dimensional grid was made because a Cartesian plane can represent more complex real-world types, such as patient health along the dimensions of heart rate, x-ray results, etc. The grid has the important concept of distance representing similarity, just as types in real-life may be similar along multiple dimensions.

The goal of the playing agent is to reach the goal as quickly as possible. Because the playing agent knows their current position, the goal position, and the shape of the board, they can deterministically calculate the optimal path using a Markov decision process (MDP). In structuring the MDP for $r$ rows and $c$ columns, we assign a positive reward of $r\*c$ to the goal state and a negative penalty of $-1$ to every other state. From any state, the possible actions are ${Up, Down, Left, Right}$, and only legal actions are defined for each state (meaning that the agent will not attempt to go off the board).

We can solve this MDP using value iteration to find $V$, where $V(s)$ is the expected value of performing the optimal policy starting at state $s$, and $Q$, where $Q(s,a)$ is the expected value of choosing action $a$ in state $s$. Because this game is deterministic and must terminate after at most $r\*c$ steps ($r+c$ steps assuming no backtracking), we can use the finite horizon value iteration algorithms to solve this MDP. The algorithm to do so is as follows:

For each iteration until our horizon

For each state

For each legal action from this state

***Equation here***

By solving this MDP, the playing agent can determine the optimal policy from any given state. If desired, we can also include the element of irrationality in the system by not always choosing the optimal action:

***Equation here***

*Type Estimates with a Single Observer*