university of groningen

faculty of mathematics and natural sciences

Design of Multi-Agent Systems

Violence FTW

Authors:
Maikel Withagen
Robin Kramer
Steven Bosch

Supervisor:
Dr. Bart Verheij

September 24, 2015

Contents

1 Problem

The civil war in Bosnia and Herzegovina from 1992 to 1995, the Ferguson unrest in 2014, and more recently the civil war in the Ukrain; Civil violence has been an issue in many countries for many years. There is no single cause for these riots and wars. Some were the result of cultural differences, some because of the feeling of being treated injustly, and others came about due to political reasons. “Each war is as different as the society producing it", and understanding the reasons for these riots to come about is of upmost importance (Ref5Goh), but also how the authorities respond to these situations should be the focus for investigation.

Much research has been done to understand how these riots emerge, including simulations of riots using a game theoretic approach (Ref8Goh) and social networks (Ref10Goh) among others. These investigations have shown that the behavior at a macroscopic level, that is the behavior of an entire group, is the result of the microscopic level, that is the behavior of the individual. In a more recent study, Goh and colleagues (Goh, Quek, Tan & Abbass, 2006) also studied how macroscopic behavior emerged using a game theoretic approach in a simulation. The most important issue the authors adressed was how experiences changed the behavior of the individual and how this learning affected macroscopic behavior.

Goh included many different interactions in his simulation, such as a probability that civilians turned to active protesters, how jail time affected rehabilitation of arrested protesters, and how the amount and types of people affected individual decisions. The focus was thus mainly on the civilian and the protesters. What the authors failed to focus on was how experience affected the learning rate. In many reïnforcement learning algorithms, such as Q-learning or Temporal Difference learning acknowledge that many previous experiences ensure a stability in behavior, resulting in a decreased influence of newer experiences, (Watkins & Dayan, 1992: Q-Learning, technical notes; REF TD). Many argue that this decrease in learning rate also occurs in humans (REF, REF). In other words, more experienced humans are less affected by new experiences compared to less experiened people. It is very well possible that the ‘correct’ choice may result in a negative outcome in a particular situation. The experienced person will be resilient to such an ‘accidental’ outcome, whereas a less experienced person may be affected more severely by that negative outcome. The latter person may therefore change to a less favourable course of action in a future occurence of said situation. This may result in frailty of the group’s dynamics, ultimately leading to a drop in performance or even a loss. On the other hand, if a less experienced person had some accidental successes with an 'incorrect' action, it is easier to learn that 'correct' actions are more profitable. Individual experience may thus have a great impact on group performance, but so far, this has not received much attention.

The current paper focusses on how experience will influence cop behavior. By using a simulation, we were able to compare how experienced cop agents, not experienced cop agents and a mix thereof perform against a group of trained hostile agents. In the followings sections, the simulation and, in turn, data acquisition will be descibed in more detail. Following the results of the different simulation will be presented. The paper concludes with a discussion of the implications of the results, and some shortcomings to this paper.

2 Method

The simulation consists three types of agents: cops, hostiles and civilians. The goal of the cops is to keep (civil and cop) casualties as low as possible. The hostiles on the other hand have only one goal. Their only goal is to kill as many civilians as possible, to cause mayhem and despair. Civilians have no particular function, but are subject to the actions taken by cops and hostiles. The simulation takes place in a NxN 2D matrix. In every box of the matrix 20 agents can reside. The agents that reside in this box can only see the 19 other agents in that box. This allows for easier computation of interactions with visible agents.

Goals and Actions The cops goal is to keep casualties as low as possible. To achieve this, the cop must choose one of two actions: shooting a hostile, with a risk of killing a civilian, or saving a civilian, with the risk of being killed by a hostile. Hostiles also have two options. They can either choose to shoot at the civilians with a chance of hitting an agent, or they can shoot at the cops for self defence, with a chance of hitting a civilian. All the cops and hostiles, during every epoch, can perform one of these actions. Based on the outcome of the agent's individual action, a reward is given to this individual agent. This reward can be either positive or negative. If a cop kills a civilian, for instance, a negative reward will be given, whilst when he has killed a hostile, a positive reward will be given.

		Action 1	Action 2
Size Team 1 $>$ Team 2	Many civilians	Ux	Uy
	Few civilians	Ur	Ut
Size Team 1 $<$ Team 2	Many civilians	Ua	Ub
	Few civilians	Uc	Ud

Table 2.1: The different scenarios an agent can encounter and fictional success values for each action.

The reward is used to update the utility of the particular actions in particular circumstances. These circumstances are shown in Table 2.1. The agents must thus decide, based on the amount of team members, opponents and civilians what action would be most successful. In the beginning of the simulation, all the utilities are set to equal values, such that no biases may exist. Reward of an action can be calculated as following for cops:

$$R = (Kills + Saves - Deaths)/(Kills + Saves + Deaths)$$

in which $Kills$ are the amount of killed hostiles, $Saves$ the amount of saved civilians, and $Deaths$ the amount of killed cops. Based on these calculations the reward will lie between -1 (only deaths) and 1 (only kills and saves). For hostiles the reward function is as follows:

$$R = (Kills - Deaths)/(Kills + Deaths)$$

, in which $Kills$ is the amount of killed civilians and cops, and $Deaths$ the amount of killed fellow hostiles.

As the Table 2.1 also shows, there is no situation in which the teams are of equal size. It is assumed that the agents have no perfect knowledge of the area, and, therefore, must decide which team has the overhand. This decision is made according to the following stochastic function:

$$functionthatdetermineswhichteamhastheoverhand.$$

After each epoch, the agents can decide to move to a neighboring box in the matrix. If the agent decides to move it will go to the place, in which the action with the highest utility will be applied. In other words, the agents will move to where he thrives best.

Learning Q-learning implements a decrese in learning rate by adjusting the learning rate variable λ over time (REF). In temporal difference learning this is achieved by storing memories (of actions and results) in time (REF). Over time these memories will decay (less often experienced ones quicker than more experienced ones) allowing a flexible learning of the best actions/strategies. For this simulation, the learning rate variable will be implemented.

Scenarios Overtal/Ondertal, veel civilians/weinig civilians.

3 Reflection