ABSTRACT
We modelled the conflict situation using a Markov game on various complex networks and investigated the emergence of conventions for conflict resolutions in agent networks with various structures through pairwise reinforcement learning. We found the network structure strongly affected their emergence and the agents could sometimes learn no conventions although they could learn locally consistent actions for resolutions.

Categories and Subject Descriptors
I.2.11 [Artificial Intelligence]: Distributed Artificial Intelligence—Multiagent systems

General Terms
Experimentation

Keywords
Conventions; Norms; Conflict resolution; Social networks

1. INTRODUCTION
Conflict resolution between agents often incur high cost due to their sophisticated reasoning with large numbers of communications. One facilitation of coordination and conflict resolution is to provide or evolve social norms and conventions that all agents are expected or learn to follow. This can significantly reduce both computational and communication costs in coordination and conflict resolutions by regulating coordination behaviors.

We introduce a modified narrow road (MNR) game [3] to represent such a conflict situation where, to resolve conflict, at least one agent is forced to follow a strategy that is superficially unacceptable, so it is not likely to select it at first. In addition, the conflict situation still remains if the agents fail to resolve it. Then, agents individually learn efficient strategies to resolve the conflict situation, where efficient strategies mean that agents can resolve conflicts with fewer actions, and thus, minimize the expected penalty. We assumed that a social convention would emerge if agents had identified efficient strategies for any game adversaries.

Our main purpose is to investigate the static/dynamic characteristics and stability of emergent conventions in the MNR game by varying two perspectives: payoff matrices representing the agents’ attitudes to conflict situations, i.e., abstractly representing how the agents act, and network structures in agent societies, called agent networks, such as Barabasi-Albert (BA) model[1] and connecting nearest neighbor (CNN) model[4] networks as well as fully connected networks (FCN), because real-world agents interact with one another according to some (physical and virtual) constraints. We experimentally show how agent attitudes and underlying network structures affect on the emergence of conventions.

2. EMERGENCE OF CONVENTIONS
An agent network for the set of agents $A$ is denoted by graph $(A, E)$, where $E$ is the set of edges. The edge between agents $i$ and $j$ is denoted by $e_{ij}$ and $j$ is called the neighbor of $i$ (and vice versa). The MNR game corresponds to a situation in which two car agents encounter each other at a narrow road (the details are described in [3]). This is a two-player Markov game [2] as shown in Fig. 1. Their rewards are denoted by a payoff matrix such as

$$(M1) \text{ Selfish} \quad (M2) \text{ Self-centered}$$

$$
\begin{pmatrix}
  p & s \\
  -5 & 3
\end{pmatrix} \quad 
\begin{pmatrix}
  p & s \\
  -5 & 3
\end{pmatrix}
$$

where the agents take one of two actions, i.e., $p$ (proceed) or $s$ (stay). Neighboring agents $i$ and $j$ play the MNR game. We investigated how agents learned the conventions for MNR games by reinforcement learning and how their society

Figure 1: State transitions in MNR game
began more efficient as a result of emergent behaviors.

If two agents take joint action \((p, p)\) or \((s, s)\), they cannot resolve the conflict (i.e., the game does not end) and they move on the second round of the MNR game with the same adversary as shown in Fig. 1, where \(S = W_0\) and \(T\) are the start and terminal states and \(W_k\) is the state of the \(k + 1\)-th round of the MNR game. Therefore, agents have already come to a standstill \(k\) times. When a convention emerges, all pair of agents take joint action \((p, s)\) or \((p, s)\) according to their sides.

\(N_g (N_a)\) is a positive integer \(g\) (i.e., pairs of agents) are selected and start the game at every tick, which is time unit. They take actions using \(\varepsilon\)-greedy strategy based on the results from reinforcement learning whose states are specified by both the number of the round in a game and what side of the game they are on left (\(L\)) or right (\(R\)) of the road. We can say that conventions are common policies learned in an agent society.

To describe the agents' policies at \(W_0\), we denote the strategy pair for each side by \(Lm_{L1}Rm_{R1}\), where \(m_L \) and \(m_R\) corresponds to the preferred actions, \(p\) or \(s\), at \(W_0\) on the left and right sides. For example, \(LpRs\) means that actions \(p\) on the left side and \(s\) on the right side are prefered. The agent network is considered to have learned the convention if the ratio of agents having strategy pair \(LpRs\) (or \(LsRp\)) is more than \(1 - T_c\), where \(0 \leq T_c \ll 1\), because almost all pairs of agents can resolve conflicts fairly in a single round of the MNR game. Note that we specially focus on \(W_0\), because our primary concern is to effectively resolve conflicts.

After a convention emerges, if the agents on the left (right) side proceed first in state \(W_0\), it is called left-priority convention (left-priority convention). We also call the prior (non-prior) side when agents at this (another) side proceed first.

3. EXPERIMENT

We conducted a number of experiments to investigate what effects payoff matrices and network structures would have on the emergence of conventions and/or on the performance of agent societies. In this experiment, we set \(|A| = 10,000\), \(N_g = 100\), \(\varepsilon = 0.05\), and \(T_c = 0.1\).

![Figure 2: Performance in BA and CNN networks](image)

**Figure 2: Performance in BA and CNN networks**

Figure 2 is a graph of the numbers of rounds per tick in agent networks, i.e., the performance of conflict resolution, in complex networks CNN and BA\(n\), where BA\(n\) means that BA networks where a new node (agent) is added with \(n\) edges [1]. It shows that a convergence almost emerged except the BA2 networks. We can also observe small hills around 80,000 when the network was BA10 (and BA5 although its hill was quite small). Note that the average rounds of the game is one if no conflicts occurred. However, since \(\varepsilon = 0.05\), they converged to slightly higher than 100 (= \(N_g\)).

We examined how many agents identified the best action \(p\) or \(s\) by looking at their Q-values at \(W_0\) to analyze these phenomena. We examined fully-connected networks (FCN), BA model networks (BA networks) [1], and CNN networks [4] but we will present the result when the network is BA5 with self-centered agents. Because which side had priority depended on each trial of the experiments, we used the terms prior or non-prior side instead of left or right side. We denote the number of agents that prefer action \(a\) when they are on the prior (non-prior) side in state \(W_k\) as \(N_{x(a)}^g\) (or \(N_{x(p)}^g(a)\)), where \(a = p\) or \(s\).

![Figure 3: Strategies selected in BA5 with self-centered agents](image)

**Figure 3: Strategies selected in BA5 with self-centered agents**

Finally, we want to point out that the alternation of conventions in self-centered agents cannot be observed in one-shot games. Thus, it is important to hold Markov games to analyze conflict situations.

4. REFERENCES


