Robotics and Autonomous Systems 61 (2013) 661–666

Contents lists available at SciVerse ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

Coordination of communication in robot teams by reinforcement learning

Darío Maravall [a], Javier de Lope [a,b], Raúl Domínguez [a,c,*]

[a] Computational Cognitive Robotics Group, Department of Artificial Intelligence, Universidad Politécnica de Madrid, Spain

[b] Department Applied Intelligent Systems, Universidad Politécnica de Madrid, Spain

[c] Center for Automation and Robotics, Universidad Politécnica de Madrid - Spanish National Council for Scientific Research, Spain

ARTICLE INFO

Article history:
Available online 4 March 2013

Keywords:
Multi-agent systems
Multi-robot systems
Dynamics of artificial languages
Computational semiotics
Reinforcement learning
Self-collective coordination
Language games
Signaling games

ABSTRACT

In multi-agent systems, the study of language and communication is an active field of research. In this paper we present the application of Reinforcement Learning (RL) to the self-emergence of a common lexicon in robot teams. By modeling the vocabulary or lexicon of each agent as an association matrix or look-up table that maps the meanings (i.e. the objects encountered by the robots or the states of the environment itself) into symbols or signals we check whether it is possible for the robot team to converge in an autonomous, decentralized way to a common lexicon by means of RL, so that the communication efficiency of the entire robot team is optimal. We have conducted several experiments aimed at testing whether it is possible to converge with RL to an optimal *Saussurean* Communication System. We have organized our experiments alongside two main lines: first, we have investigated the effect of the team size centered on teams of moderated size in the order of 5 and 10 individuals, typical of multi-robot systems. Second, and foremost, we have also investigated the effect of the lexicon size on the convergence results. To analyze the convergence of the robot team we have defined the team's consensus when all the robots (i.e. 100% of the population) share the same association matrix or lexicon. As a general conclusion we have shown that RL allows the convergence to lexicon consensus in a population of autonomous agents.

1. Introduction

In a multi-robot system obtaining a common lexicon or vocabulary is a basic step towards an efficient performance of the whole system [1]. In this paper we present the application of Reinforcement Learning (RL) to the emergence of a common lexicon in a team of autonomous robots. We model the vocabulary or lexicon of each robot as an association matrix or look-up-table that maps the *meanings* (i.e. the objects, states of the environment and self-states) into *symbols* or *signals*. According to a long and well-established line of thought culminating with the work of Ferdinand de Saussure [2] and Charles S. Peirce [3], the pioneer of Semiotics, the association of the symbols of a language to their meanings are (1) arbitrary and (2) conventional.

In this paper we use arbitrarily (in fact, randomly) initialized association matrices, for each robot and through a dynamic process based on *communicative* or *linguistic interactions* implemented by means of the Reinforcement Learning paradigm, the team converges to an optimum consensus state.[1]

* Corresponding author at: Computational Cognitive Robotics Group, Department of Artificial Intelligence, Universidad Politécnica de Madrid, Spain.

E-mail addresses: dmaravall@fi.upm.es (D. Maravall), javier.delope@upm.es (J. de Lope), r.dominguez@alumnos.upm.es (R. Domínguez).

[1] Some authors, inspired in the ideas of Ludwig Wittgenstein, have called them *language games* [4–6], although we believe a more founded denomination should be *communication* or *signaling games* as defined by David K. Lewis [7].

0921-8890/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.robot.2012.09.016

2. Formal definitions

2.1. Multi-robot communication system

We define a Communication System, *CS*, in a team of robots as the triple:

$$CS \triangleq \langle M, \Sigma, A_i \rangle \tag{1}$$

where $M = \{m_1, \ldots, m_p\}$ is the set of meanings (i.e. the objects or states in the environment that can be of relevance for communication in the team of robots), $\Sigma = \{s_1, \ldots, s_n\}$ is the set of symbols or signals used by the robots in their communication acts and which represent the actual meanings, $A_i\ (i = 1, \ldots, N)$ are the association matrices of the robots defining their specific associations between meanings and symbols:

$$A_i = \left(a_{rj}\right)_i ; \quad i = 1, \ldots, N \text{ agents} \tag{2}$$

in which the entries a_{rj} of the matrix A are non negative real numbers such that $0 \leq a_{rj} \leq 1, (r = 1, \ldots, p; j = 1, \ldots, n)$. These entries a_{rj} give the strength of the association of meaning m_r to symbol s_j; such that $a_{rj} = 0$ indicates no association at all and $a_{rj} = 1$ indicates total association. Note that these quantitative associations have a deterministic and non probabilistic nature so that the associations between meanings and symbols are based on the maximum principle, which means that the maximum value of the entries in a row (column) gives the valid association. An ideal, optimum association matrix is purely binary (the entries are either 0 or 1) and also have the additional restriction of having in each row only one 1 (i.e. no synonyms are allowed) and having a unique 1 in each column, too (no homonyms are allowed).

2.2. Optimal communication efficiency

The necessary and sufficient condition for an optimal communication efficiency between two generic robots A_s and A_r is that the optimal association matrix of the receiver robot must be equal to the optimal association matrix of the sender robot:

$$A_s = A_r \tag{3}$$

where A_s is the optimal association matrix of the robot acting as the sender and A_r is the optimal association matrix of the robot acting as the receiver.

2.3. Optimal multi-robot communication system

Given N robots with respective association matrices: $A_1, A_2, \dots, A_N$. The necessary and sufficient condition for an optimal global communication efficiency is that all the robots must share the same optimal permutation matrix $A = A_1 = A_2 = \cdots = A_N$.

According to the semiotic tradition, the associations of the symbols of a lexicon to its meanings are arbitrary and conventional. The arbitrariness of the association of a symbol (the signifier in Saussure's jargon) to its meaning (the signified) means that the entries a_{rj} of each robot's association matrix are arbitrarily assigned. The conventional nature of these associations means that all the robots of the population must have the same ideal, optimum permutation association matrix in order to attain an optimum global communication efficiency (we call *Saussurean* such an optimal communication system). This situation is called lexicon or vocabulary consensus and is a hard multi-robot coordination problem. In the sequel we present our experimental work showing that Reinforcement Learning can solve this multi-robot coordination problem.

3. Dynamics of convergence to lexical consensus by means of reinforcement learning

3.1. General description of the process

For a team of N robots, each of them with its own lexical matrix A_i $(i = 1, 2, \dots, N)$ the way to converge to lexical consensus by means of reinforcement learning is based on the following procedure (see pseudo code in Fig. 1). First, a sequence of what we call language games rounds is performed until the team converges to an optimal communication system or lexical consensus in which all the robots use the same optimal permutation association matrix. In each language game round, all the possible pairwise communication acts are performed (see pseudo code in Fig. 2). Each communication act taking place between two robots with corresponding lexical matrices $\{A_i, A_j\}$ $(i = 1, 2, \dots, N; j = 1, 2, \dots, N; i \neq j)$ proceeds as follows: one of the two robots is selected at random as the speaker (sender) and the other as the listener (receiver). The speaker sends all the possible meanings according to its association matrix. If the meaning decoded by the listener coincides with the speaker's meaning then a success has happened. A failure happens when both meanings differ. After a success the corresponding coefficients of the association matrices in both robots are increased and the competing association coefficients (i.e. a row for the speaker and a column for the listener) are updated in the opposite direction. This additional updating is known as *lateral inhibition* and it is a key element for the convergence process. Similarly, the coefficients involved in a failure are decreased in both robots.

```
for k = 1, 2, ..., max rounds do
  Execute all the possible communication acts
  Compute the communicative efficiency of the robot team EC(k)
  if EC(k) = Max in three consecutive rounds then
    Break
  end if
end for
```

Fig. 1. Pseudo code of the reinforcement learning-based lexical coordination procedure.

```
Assign randomly the sender/receiver roles
for k = 1, 2, ..., number of meanings (randomly ordered) do
  Send the meaning m_k according to the sender's association matrix
  Decode the received symbol s_k according to the receiver's association
  matrix
  Update both matrices depending on the communication result
end for
```

Fig. 2. Pseudo code of a communication act.

The ultimate goal is that after the execution of all the language games rounds the robot team converges to an optimal communication system in which all the robots use the same optimal permutation matrix (optimal *Saussurean* solution).

3.2. Algorithms for the updating of the association matrices

We have applied two different algorithms for the updating of the coefficients of the association matrices: (a) an Ant Colony Optimization-based algorithm, or ACO-like for short, and (b) the incremental algorithm.

In the ACO-like algorithm the coefficients of the association matrix are updated as follows:

$$a_{ij}(k+1) = \rho\, a_{ij}(k) + (1-\rho)\,\beta(k) \quad 0 \leq \rho \leq 1$$
$$\beta(k) = \begin{cases} 1 & \text{if reward/success} \\ 0 & \text{if punish/fail} \end{cases} \tag{4}$$

in which ρ is a critical parameter which has to be carefully selected [8].

In the incremental algorithm [6] the coefficients are updated as follows:

$$a_{ij}(k+1) = a_{ij}(k) + \Delta\,\beta(k) \quad 0 \leq \Delta \leq 1$$
$$\beta(k) = \begin{cases} 1 & \text{if reward/success} \\ 0 \text{ or } -1 & \text{if punish/fail} \end{cases} \tag{5}$$

where the parameter Δ is also critical. In this case, the coefficients are normalized after the update.

4. Experimental results

As we are mainly interested in on-line real-time implementation of lexical coordination in physical multi-robot systems, our simulation experimental work has revolved around moderate team sizes typical of multi-robot systems, in the order of 10 or less robots. For them we have experimented with different lexicon dimensions and complexity, as regards speed and stability convergence.

We have conducted series of 600 experiments with the lexicon dimensions from 2×2 to 7×7 (meanings $\times$ signals). In all these cases we have centered our efforts on studying the convergence to a common *Saussurean* lexicon for teams of 5 and 10 robots.

Once evaluated the convergence dynamics, we conclude that the considered coordination problem can be solved by reinforcement learning techniques with a sufficient success rate for the cases of lexicons of 2 meanings and 2 signals and that the RL method is particularly efficient in the case of teams of 5 robots.

As the lexicon complexity or the team size increases the rate of success of convergence to a *Saussurean* communication system decreases due to the fast growth of the problem's search space.