Decentralized multi-agent patrolling strategies using global idleness estimation

Mehdi Othmani-Guibourg, Amal El Fallah-Seghrouchni, Jean-Loup Farges

► To cite this version:

Mehdi Othmani-Guibourg, Amal El Fallah-Seghrouchni, Jean-Loup Farges. Decentralized multi-agent patrolling strategies using global idleness estimation. PRIMA 2018: The 21st International Conference on Principles and Practice of Multi-Agent Systems, Oct 2018, TOKYO, Japan. hal-02335231

HAL Id: hal-02335231

<https://hal.science/hal-02335231>

Submitted on 28 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Decentralized multi-agent patrolling strategies using global idleness estimation

Mehdi Othmani-Guibourg^{1,2}, Amal El Fallah-Seghrouchni², and Jean-Loup Farges¹

¹ ONERA, Toulouse, France {first name.name}@onera.fr

² Sorbonne Université - Faculté des Sciences, CNRS, UMR 7606, LIP6, F-75005, Paris, France
{mehdi.othmani; amal.elfallah}@lip6.fr

Abstract. This paper presents preliminary results in the challenge of developing decentralised strategies approaching the performances of centralised ones. Indeed, the latter are better than the former due to centralisation of information. The approach studied here involves the estimation of node idlenesses derived from the paths of all agents, also known as real idlenesses, on the basis of those derived from the path of each agent considered alone, also known as individual idlenesses. This relation between real and individual idlenesses is learnt using traces of execution of a centralised strategy by optimising an error criterion. The strategy thereupon, uses online the learnt relation and is assessed according to certain evaluation criteria. The results indicate that such a relation between perceived and real idlenesses is not a function, leading to large values of the fitting criterion. Finally, the assessment of the strategy shows that performances are good in terms of mean interval but unsatisfactory in terms of quadratic mean interval.

Keywords: Multi-agent systems · Multi-agent patrolling · Artificial Neural Networks

1 Introduction

The patrol task is well-suited for being shared in space and time by several agents. There are a wide variety of problems that may be reformulated as a particular multi-agent patrol task. As a concrete example, the monitoring of an area by a swarm of drones does face the problem of coordinating them to patrol that area in order to detect certain events. A feature of multi-agent patrolling (MAP) is the difficulty to derive analytic results from its equations. Thereby it appears that the only method enabling to predict its behaviour is to simulate the local interactions of its components. Thus, the quality of a patrolling strategy is evaluated in simulation and it is consensual that a good strategy is one that minimises the time lag between two passages on the same place and for all places.

Different types of strategies were proposed, however, few works concentrate on the problematic of using Artificial Neural Networks (ANN) for the MAP [2] [4] [7]. *We propose and evaluate new strategies using not only machine learning models, in particular an ANN based on the rectifier linear ReLU, but also a multi-dimensional linear model. In this way, new strategies embedding these models are introduced.*

The Section 2 presents the background useful to understand proposed developments: the MAP and ANN types used by the strategies. Then, Section 3 describes new strategies based on idleness estimation by ANN. In Section 4 these strategies are analysed. Finally, Section 5 draws some conclusions and indicates directions for further works.

2 Background

2.1 Multi-agent patrolling

The MAP model consists of a society of agents noted $\mathbf{A}$, able to move in a graph noted $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ representing a discretisation of the area to patrol. $\mathbf{V} = \{1, \dots, N\}$ is the set of nodes standing for the places to visit, and $\mathbf{E}$ is the set of edges connecting them. At each edge corresponds a *transit time* representing its travel time. At each node is associated a dynamic variable named *idleness*, indicating the time elapsed since it has not been visited by any agent [3]. The vector of idlenesses of all nodes at time t is noted $I_t(v)$ and the idleness of a node v , $I_t(v)$. At the beginning of a patrolling, agents are positioned on nodes and all idlenesses are set to 0. Finally, each time an agent arrives at a node v , it shall decide, among the edges connecting v , the next edge to travel.

A *strategy* of agent is an information processing method allowing each agent to take a decision each time it arrives at a node. Whatever the strategy considered, each agent intends actions based on its knowledge regarding idlenesses of nodes. Indeed, agents make idleness estimates that can be produced assuming different hypotheses:

- *individual idleness*: each agent considers only its own visits to update its estimated node idleness. It corresponds to the case where communication between agents is not possible. In the case of a mission with only one agent, *individual idleness* corresponds to *real idleness*, also called *global idleness*.
- *shared idleness*: all agents consider visits of all agents to reset estimated node idleness. In the case of perfect instantaneous communication between agents, or a mission with only one agent, *shared idleness* corresponds to *real idleness*, also called *global idleness*.

Among the wide family of strategies [1], two regarded as representative strategies are relevant here: *Conscientious Reactive* (CR) and *Heuristic Pathfinder Cognitive Coordinated* (HPCC). CR selects the next node to visit as the one with the highest individual idleness in its neighbourhood. HPCC is based on a perfect communication between agents: shared idlenesses are estimated by a coordinator on the basis of all paths of agents. Two methods are used: the first one called *Heuristic* selects the next node to visit, and the second one called *Pathfinder* chooses the path to go there [1] [5].

Evaluation criteria relevant to establish aggregation measures based on *interval between visits* for a node are the *Mean Interval* (MI) and the *Quadratic Mean Interval* (QMI). In order to better evaluate the contribution of each agent when the population size varies, these criteria are normalised by multiplying values by the number of agents.

2.2 Artificial neural networks

ANNs are a special kind of *machine learning* models. Among the large variety of ANN, single layer and **multi-layer perceptrons** [8] [6] are composed with one or several stacked layers of *neurons*, each one corresponding to a function that maps the outputs of the previous layer with the output of the current layer. Several kinds of functions are used. For example, the *identity*, the *logistic sigmoid* and the *linear rectifier* (ReLU), $f(z) = \max(0, z)$. Each neuron computes a weighted and biased sum of the previous layer's outputs, which is finally passed through its function, making up thereby the neuron's output. When the functions are non-linear, they provide a basis for developing an approximation of the function to be learned.

Networks are generally optimised using *gradient-descent-based methods* by minimising a *cost function* representing the difference between the output of the network and its desired value [8] [6]. A quite common cost is the *Mean Squared Error* (MSE).

3 Strategies based on idleness estimation

This section presents the design of the three strategies using an idleness estimation called *Heuristic Pathfinder Mean Predictor* (HPMP), *Heuristic Pathfinder Linear Predictor* (HPLP) and *Heuristic Pathfinder ReLU Predictor* (HPRP).

3.1 Formal definition

Estimator-based strategies The three strategies use an estimator based on a trained statistical model noted $m(.,.)$: the decision-making process is carried out first by computing an estimate of the *global idleness* from the trained model, then by making the decision regarding the next node to visit with respect to this estimate. In our context, a temporal series representing the successive idlenesses each time an agent stands upon a node, the latter will be called an *idleness flow*.

Let I_t^a and $\hat{I}_t$ being the vectors of individual idlenesses of the agent a and the corresponding estimated global idlenesses, respectively, at the time t . Then, given that $\forall t \in \mathbf{T}, \forall a \in \mathbf{A}, \forall I_t^a \in \mathbb{R}^N$, $\hat{I}_t$ is defined such as:

$$\hat{I}_t = \min(\max(m(I_t^a, \theta), 0), I_t^a) \quad (1)$$

where θ is the set of the statistical model's parameters, and min and max are component-wise functions ensuring that whatever the output of the model, the estimation of global idleness is positive and lower than the individual idleness.

For any agent, such an estimator will output estimates of the global idlenesses, called *estimated idlenesses*, from the current individual idlenesses fed as input. All the agents embed the same estimator, i.e. the same trained model. This strategy can be thought of as a reactive strategy using an artefact for estimating missing information regarding the area to patrol, and taking into account the idleness of nodes and thereby implicitly the agents' positions.

For a given scenario, the model learns to predict the global idleness vector corresponding to a current agent's individual idleness vector. Then, agents applies to the

estimated global idlenesses the two methods described in the **Section 2**, namely the Heuristic method to select the next node to visit and the Pathfinder method to choose the path to go there.

However, the relation between individual idlenesses and global idlenesses may not be a function. The following theorem presents conditions under which this relation is not a function.

Theorem 1. *Let $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ be a graph, let $\mathbf{A}$ be a society of agents and consider two runs of a given strategy, arbitrarily named first and second run. If:*

- *in initial state, a node $u \in \mathbf{V}$ is occupied with an agent $a_1 \in \mathbf{A}$ for the first run and with an agent $a_2 \in \mathbf{A}$ for the second run, a_1 and a_2 may have the same agent identifier, and*
- *a next node $v \in \mathbf{V}$ is selected by the strategy for the agent a_1 in the first run and for the agent a_2 in the second run and*
- *for the first run, it exist $w \in \mathbf{V}$, $w \neq v$ that is occupied by an agent $a_3 \in \mathbf{A}$, $a_3 \neq a_1$, $a_3 \neq a_2$ or which has already been reached by a_3 when, at time t , a_1 arrives at v and*
- *for the second run, no agent has reached w , when at time t , a_2 arrives at v ,*

then the relation between, the individual idleness I_t^a and the global idleness I_t is not a function.

Proof. At time t , the individual idlenesses for a_1 in the first run and for a_2 in the second run are equal: both have $I_t^a(j) = t, \forall j \neq v$ and $I_t^a(v) = 0$. For the first run $I_t(w)$, the global idleness of w , is equal to 0 if a_3 occupies it, or equal to $t - \tau < t$, where $\tau > 0$ is the travel time of agent a_3 from its initial position to w , otherwise. For the second run $I_t(w) = t$. Thus, to the same individual idleness corresponds two different values of the global idleness. Hence, the relation between the individual idleness I_t^a and the global idleness I_t is not a function.

Models As indicated by the **Eq. 1**, for all the models studied here, the input and output both of dimension N , stand for the vector of individual idlenesses and the vector of estimated idlenesses, respectively.

First, the *mean model* consists of a model which estimates for each node, the global idleness as being the average of all global idlenesses of this node over all the global idleness flows. With such a model noted *Mean*, an agent $a \in \mathbf{A}$ carries out the estimation of the global idlenesses at $t \in \mathbf{T}$ as following:

$$\exists \theta = \{B \in \mathcal{M}_{N \times 1}(\mathbb{R})\} : m(I_t^a, \theta) = \text{Mean}(I_t^a, \theta) = B \quad (2)$$

When such a model is used as global idleness estimator, the corresponding strategy is called Heuristic Pathfinder Mean Predictor (HPMP).

When the estimator corresponds to a linear model noted *Lin* or *Linear*, the strategy is termed Heuristic Pathfinder Linear Predictor HPLP. With such a model an agent $a \in \mathbf{A}$ carries out the estimation of the global idlenesses at $t \in \mathbf{T}$ as follows:

$$\exists \theta = \{W \in \mathcal{M}_N(\mathbb{R})\} : m(I_t^a, \theta) = \text{Lin}(I_t^a, \theta) = W \cdot I_t^a \quad (3)$$

with W being the model's weight matrix. Training such a model corresponds to figure out the W minimising a certain distance between $m(I_t^a, \theta)$ and I_t .

Finally, an MLP composed with H hidden ReLU layers as described in the **Section 2**, is termed *ReLU model*, while its corresponding strategy is called Heuristic Pathfinder ReLU Predictor (HPLP). With such a model noted MLP_{ReLU}^H , an agent $a \in \mathbf{A}$ carries out the estimation of the global idleness at $t \in \mathbf{T}$ as following:

$$\begin{aligned} \exists \theta = \{W_h, W_{out} \in \mathcal{M}_N(\mathbb{R}) : h \in [1, H]\} : \\ m(I_t^a, \theta) &= MLP_{ReLU}^H(I_t^a, \theta) \\ &= W_{out} \cdot ReLU(W_H \cdot ReLU(W_{H-1} \cdot ReLU(\dots \\ &\quad W_2 \cdot ReLU(W_1 \cdot I_t^a) \dots))) \end{aligned} \quad (4)$$

with *ReLU* being the element-wise ReLU activation, and $\forall h \in [1, H]$, W_h the weight matrix of the layer h .

4 Experiments and results

4.1 Scenarios and training

Fig. 1: Graphs used during assessment.

Three different graphs were selected to evaluate the strategies: the maps *Islands*, *Grid* and *A*, as shown in the **Figure 1** [5]. For each map we tested the strategies CR, HPCC and the idleness-predictor-based strategies were trained from HPCC's simulation and tested. The tests were performed over population sizes of 5, 10, 15 and 25 agents and for each size we selected 100 *random starts*, also called *runs*. For each start, each strategy was tested over 3000 periods and, in average, an agent visits 650 nodes during one execution of 3000 periods. In doing so, the sequences used to train the models have approximately a length of 650 idleness vectors. For each scenario we trained 8 statistical models by minimising the MSE: a mean-based model, a linear model, three Multi-layer Perceptron (MLP) with sigmoid units and three different artificial neural networks with rectifier linear units (ReLU): an architecture with only one ReLU layer simply termed *ReLU*, another one with one hidden ReLU layer and the output layer being also a ReLU layer, termed *ReLU Output* (ReLUO), and finally an MLP with ReLU activation termed *ReLU MLP*. The data base was divided into a training base and a validation base with

Fig. 2: Normalised MI of the evaluated strategies in ordinate for the three maps w.r.t. the population sizes of agents in abscissa.

80% of data in the training base. The training results indicates that the lowest MSE are obtained for a single layer of linear or ReLU neurons. Strategies based both on these networks and on the mean were selected for assessment.

4.2 Performance results

To evaluate their performances, the studied strategies were tested and compared with CR and HPCC using normalised MI and QMI as evaluation criteria.

Fig. 2 show all the results for the topologies Islands, A and Grid for the normalised MI. Not surprisingly, HPCC always outperformed all the others strategies on all the maps and for all the population sizes of agents. First, except for the map Grid, HPLP overwhelmingly outperforms the reactive strategy CR, while on the map Grid it is slightly better than CR. Unlike the others maps, this little difference can be explained in considering that, Grid being a topology where the nodes are uniformly distributed, the strategy CR is well adapted. However, HPLP remains better than CR on this map, except for the population size of 15 agents where they are approximately equal. Then, except for the map Islands, HPLP has always better performances over this criterion than HPMP. On that map, the performances of HPLP are approximately equal to the ones of HPMP for 5 and 10 agents. However, for 15 and 25 agents the former is worse than the latter. Results on Islands for 15 and 25 agents, on A for 5 and 15 agents and Grid for all the population sizes, seem to show that agents do not benefit from the presence of each other. Indeed, unlike HPCC which has a decreasing or stable normalised MI, it increases for HPLP. For the map Islands, HPRP is the best idleness predictor strategy over the normalised MI, except for 10 agents where it is approximately equal

Fig. 3: Normalised QMI of the evaluated strategies in ordinate for the three maps w.r.t. the population sizes of agents in abscissa.

to HPLP and HPMP. On that map, it is also slightly better than HPCC for 5 agents. For the map A, HPRP is by far the best strategy. In average it is better than HPCC of 74 periods. Finally, as previously stated while comparing HPLP and HPRP to it, HPMP is most of the time the worst idleness predictor strategy and for the map Grid it is even worse than CR of 39 periods in average. The two models trained and used as a part of the two strategies HPLP and HPRP are thereby better than HPMP.

The **Fig. 3** shows the results for the normalised QMI. Unlike the results of MI, on the three topologies the idleness predictor-based strategies are worst than HPCC and CR, the coordinated and the decentralised ones, respectively. For Islands, the HPRP is the worst strategy, while for the maps A and Grid, it is HPMP, which is the worst. For the latter maps, HPLP is incomparably better than the other two idleness predictor strategies with a difference in average with HPRP of 574 for A and 1161 periods for Grid. However, HPLP has worse performances than CR of 1522, 310 and 211 periods on the Islands, A and Grid, respectively. Finally the idleness predictor-based strategies show bad performances over the criterion QMI. QMI as quadratic mean takes better into account the difference of time intervals between the nodes and thereby measures the tendency of nodes to be equitably visited. A node with a long interval will have a little impact on MI, while it will have a large one on the QMI due to its quadratic growth.

Thereupon, the results show that good performances in average i.e. over MI are balanced by the ones of QMI. These results show the tendency of idleness predictor agents to visit a particular inferred set of nodes at the expense of the other ones.

5 Conclusion and perspectives

We proposed and evaluated new strategies for the MAP. Those strategies are based on learning the relation between individual and global idlenesses. The assessment of selected strategies based on estimations of global idlenesses using learned model indicates good results in terms of MI, but also unsatisfactory results in terms of QMI.

Theorem 1 indicates that there may be no significant expectation for approximation improvement. Data analysis methods should be applied in order to check the presence of conditions implying that the relation between individual and global idlenesses is not a function. Other future research will aim at modifying the strategies in order to improve their performance in terms of QMI. One track is to consider some randomisation process when exploiting the estimation of global idlenesses by the model. For example, defining a probability distribution with $\hat{I}_t$ as a mean and the strategy could sample in this distribution, idlenesses' estimate. The global idleness approximation problem could be further investigated with other structures of ANN and better learning algorithms. Finally, it should be noted that there is a large set of possibilities for using ANN to try to learn some information from centralised strategies that is useful for decentralised strategies. For example, nodes sequences of centralised strategies could be learned using Long Short-Term Memory ANN architectures and directly used in a decentralised strategy. This kind of approach could be compared to the approach proposed here.

References

1. Almeida A., PM Castro, TR Menezes et GL Ramalho. Combining idleness and distance to design heuristic agents for the patrolling task. In *II Brazilian Workshop in Games and Digital Entertainment*, pages 33–40, 2003.
2. David B. D'Ambrosio, Skyler Goodell, Joel Lehman, Sebastian Risi, Kenneth O. Stanley: *Multirobot Behavior Synchronization through Direct Neural Network Communication*. ICIRA (2) 2012: 603-614
3. Chevaleyre, Y.: Theoretical Analysis of the Multi-agent Patrolling Problem. In: *Proc. of the Int. Conf. On Intelligent Agent Technology*, Beijing, China, pp. 302–308 (2004)
4. Guo, Y., Parker, L., Madhavan, R.: 9 Collaborative Robots for Infrastructure Security Applications. In: *Studies in Computational Intelligence (SCI)*, April 22, v 2007, vol. 50, pp. 185–200. Springer, Heidelberg (2007)
5. Mehdi Othmani-Guibourg, Amal El Fallah-Seghrouchni, Jean-Loup Farges, Maria Potop-Butucaru. Multi-agent patrolling in dynamic environments. *2017 IEEE International Conference on Agents (ICA)*. 2017.
6. Rumelhart, D. E., Hinton, G.E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533-536 (1986).
7. D. O. Sales, D. Feitosa, F. S. Osorio and D. F. Wolf, "Multi-agent Autonomous Patrolling System Using ANN and FSM Control," *2012 Second Brazilian Conference on Critical Embedded Systems*, Campinas, 2012, pp. 48-53.
8. Werbos, P. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. PhD thesis, Harvard Univ. (1974).