How to Promote Generalisation in Evolutionary Robotics: the ProGAb Approach

Tony Pinville, Sylvain Koos, Jean-Baptiste Mouret, Stéphane Doncieux

To cite this version:


HAL Id: hal-00633928
https://hal.archives-ouvertes.fr/hal-00633928
Submitted on 19 Oct 2011

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.
How to Promote Generalisation in Evolutionary Robotics: the ProGAb Approach

Tony Pinville
pinville@isir.upmc.fr
Jean-Baptiste Mouret
mouret@isir.upmc.fr
Stéphane Doncieux
doncieux@isir.upmc.fr
ISIR, Université Pierre et Marie Curie-Paris 6, CNRS UMR 7222
4 place Jussieu, F-75252, Paris Cedex 05, France

ABSTRACT
In Evolutionary Robotics (ER), controllers are assessed in a single or a few environments. As a consequence, good performances in new different contexts are not guaranteed. While a lot of ER works deal with robustness, i.e. the ability to perform well on new contexts close to the ones used for evaluation, no current approach is able to promote broader generalisation abilities without any assumption on the new contexts. In this paper, we introduce the ProGAb approach, which is based on the standard three data sets methodology of supervised machine learning, and compare it to state-of-the-art ER methods on two simulated robotic tasks: a navigation task in a T-maze and a more complex ball-collecting task in an arena. In both applications, the ProGAb approach: (1) produced controllers with better generalisation abilities than the other methods; (2) needed two to three times fewer evaluations to discover such solutions.

Categories and Subject Descriptors
I.2.9 [Computing Methodologies]: Artificial Intelligence—Robotics; I.2.6 [Computing Methodologies]: Artificial intelligence—Learning

General Terms
Algorithms

1. INTRODUCTION
In many robotic applications, the robot has to handle a lot of situations in its daily usage, which are different from the situations used to design its controller: the controller has to show sufficient generalisation abilities (GAb), so that the robot can behave well in all the situations. For instance, if the robot has to perform a task in a room, its performance should not critically depend on the size of this room, nor on its initial position or orientation, and so on. As Evolutionary Robotics (ER) [24, 29] aims at automatically designing controllers for robots, it has to address this need for GAb of controllers: how can a controller be evaluated in only few instances of a task and also be successful in different contexts? If a single (or a few) context is used to evaluate the possible solutions, the performance value does not always reward general behaviors compared to specific ones: there is no guarantee that the best solutions are the most general. For all practical purposes, the best evolved solutions often achieve bad GAb and one can argue that the lack of a general methodology to promote the GAb of controllers curbs the application of ER to real-world robotic tasks [13].

The problem of promoting the generalisation ability has been much studied in the framework of supervised machine learning [1]. Most of the works in supervised learning dealing with generalisation are based on a so-called three data sets methodology [15], which consists in assessing and validating a model by splitting the learned contexts into three different sets: 1) a training set used to fit the model; 2) a validation set used to measure its generalization capability and to prevent overfitting; 3) a test set used afterwards to assess the GAb of the model on several contexts, which were not used during the learning. This three data sets methodology, also applied in GP [15, 30], appears to be an efficient and simple way to promote and assess the GAb of optimized solutions. A methodology to Promote the Generalisation Ability, the ProGAb...
approach, which aims at:

- finding more general controllers than state-of-the-art methods in ER;
- minimizing the number of evaluations needed to discover such solutions.

A controller which is optimally general on a given task should behave well in all the possible instances of this task. For practical purposes, this generalisation ability (GAb) is assessed on a large finite set of previously selected instances of the task that have not been used when evaluating the solutions, like the test set in supervised learning. A controller is then optimally general, if it is optimal on both training set and if it achieves similarly high performances on test set.

Promoting the GAb of controllers highlights the following conflict: 1) while the performance value on a small training set is fast to compute, it can quickly lead to over-fitted solutions with low generalisation abilities; 2) the GAb on a larger set is more informative but cannot be computed for each evaluated solution because of the computational cost.

In order to keep the number of evaluations on as low as possible, we resort to a methodology based on three major aspects: 1) all the evolved solutions are evaluated on a small training set; 2) we build during the optimization process a surrogate model which approximates the GAb on a larger set with function interpolation techniques; 3) to update this surrogate model, few controllers are assessed on this larger set while optimizing. The solutions are optimized with a Pareto-based Multi-Objective Evolutionary Algorithm (MOEA) in which two goals are defined: maximizing the performance value on the small training set and maximizing the approximated GAb on the larger set. A similar methodology has been successfully applied to the reality gap problem in [21].

The ProGAb approach has been tested on two robotic tasks performed in simulation: a navigation task in a T-maze inspired from [38] and [31] and a more complex ball-collecting task in an arena, previously introduced in [12].

2. RELATED WORK

2.1 Formalising the Generalisation Ability

Before discussing previous works dealing with the GAb of controllers, we propose a formal definition of this generalisation ability to be used in our analysis of the literature. Generalisation has been much studied in the field of supervised machine learning. Our definition is inspired from the standard three data sets methodology [15] used in supervised learning and is based likewise on three sets of contexts:

- a training set of contexts $\Omega_{\text{train}}$ is used to optimize the solutions;
- a validation set of contexts $\Omega_{\text{valid}}$ allows to estimate the GAb of some meaningful solutions;
- at last, a test set \(^1\) of contexts $\Omega_{\text{test}}$ is brought into play to assess afterhand the GAb of the best found solutions on some not yet encountered contexts.

\(^1\)The literature in machine learning sometimes reverses the meaning of “validation set” and “test set”.

In an ER experiment, a context $\omega$ is an instance of the task to solve and corresponds to a set of specific values for all the parameters of the task: initial position of the agent, size of the environment, and so on. As the training set $\Omega_{\text{train}}$ and the validation set $\Omega_{\text{valid}}$ are both used to evaluate solutions during the optimization, we also define the evaluation set $\Omega_{\text{eval}} = \Omega_{\text{train}} \cup \Omega_{\text{valid}}$, which contains all the contexts possibly used for evaluation.

Let $F$ be the fitness function defined on the task, let $x$ be an individual of the controller space $\mathcal{C}$, the generalisation ability $G$ of $x$ on a set of contexts $\Omega$ is defined as follows:

$$G(x, \Omega) = \sum_{\omega \in \Omega} F(x, \omega)$$

An individual $x$ is optimally general if it verifies the two constraints below:

\begin{align}
\forall y \in \mathcal{C}, \quad G(y, \Omega_{\text{eval}}) & \leq G(x, \Omega_{\text{eval}}) \quad \text{(C1)} \\
\frac{G(x, \Omega_{\text{eval}})}{|\Omega_{\text{eval}}|} & \approx \frac{G(x, \Omega_{\text{test}})}{|\Omega_{\text{test}}|} \quad \text{(C2)}
\end{align}

An optimally general solution has to behave at best on all the evaluation contexts (both from $\Omega_{\text{train}}$ and $\Omega_{\text{valid}}$) (C1) and to behave similarly well on the validation contexts (C2). An additional constraint is that the solutions have to be mainly assessed on the training set $\Omega_{\text{train}}$ and as infrequently as possible on the validation set $\Omega_{\text{valid}}$, although the performance on $\Omega_{\text{train}}$ and the GAb on $\Omega_{\text{valid}}$ can be conflicting.

2.2 Promoting robustness

The very first method introduced in ER to deal with the GAb of controllers consists in adding noise to the inputs of the controller [33] in order to evolve solutions which can exhibit robust behaviors to sensor uncertainties. Some other methods rely on adaptive controllers, like dynamically-rearranging neural networks [20] or plastic neural networks, which encode learning rules [14]. Such structures should adapt to environmental changes and be able to retrieve good behaviors when the context is slightly modified.

These works deal with robustness, which can be interpreted as a particular case of generalisation. It corresponds to the GAb regarding only small variations of the environment or of the controller: for instance, adding noise to the sensors or testing close initial positions. Contrariwise, generalisation does not make any assumption on the closeness from a context to another: the size of the environment can double or the starting position of the robot can radically change. Consequently, the approaches described above are not necessarily adapted to enhance the GAb of controllers.

2.3 Promoting generalisation

A review of some ER papers (table 1) indicates that, although a lot of works tackle robustness issues, only a few ones deal with the GAb of controllers in a more general sense. On the whole, the most common method consists in
evaluating the controllers with a fixed set of arbitrary evaluation contexts, often less than 10 (table 1). For instance in [12], three different initial positions are used to assess the controller for a ball-collecting task. Similarly, in [38], controllers are optimized for a navigation task in six contexts depending on the shape of the maze and the initial position of the robot. In these works, the generalisation abilities of the optimized individuals are not assessed on new contexts.

Some other methods rely on a set of evaluation contexts, which is not fixed during the optimization. For instance, in [18], the individuals are evaluated on 10 contexts, which are randomly picked up for each individual at each generation in a very large set of evaluation contexts. This evaluation set can also be co-evolved with the solutions as a subset of all the possible contexts like in [3]. The best individual found at the end of the optimization is then tested on all the possible contexts to assess its generalisation ability. Nevertheless, each context used for testing has possibly been used for evaluation, although \( \Omega_{\text{eval}} \) and \( \Omega_{\text{test}} \) should be independent to ensure good estimations of the GAb.

In [2], vision-based controllers are evolved to avoid obstacles in an arena. The individuals are evaluated on four contexts depending on the initial position of the robot (\( \Omega_{\text{eval}} \)). The individuals of the last generation are re-evaluated on four different contexts with a different configuration of obstacles (\( \Omega_{\text{test}} \)) and the best one is selected depending on its performances on these new situations. It is more adapted to truly assess the GAb of the controllers, but such an approach cannot be straightforwardly used if \( \Omega_{\text{eval}} \) is large because of high computational costs.

### 2.4 Reducing the number of evaluations

To deal with the constraint C1, a trivial approach boils down to directly optimise solutions on the whole evaluation set: each solution is evaluated on each context of \( \Omega_{\text{eval}} \). But, as the evaluation set can be very large, it is often infeasible in a reasonable computational time. The problem of assessing the GAb cannot be addressed in general without considering the need to reduce the number of evaluations.

Several methods have been proposed to lower the number of evaluations in evolutionary algorithms. The racing algorithm [5, 16] consists in sequentially evaluating candidate solutions and discarding the poor ones as soon as statistically sufficient evidence is gathered against them. The early stopping technique [7] is similar, as it ends the evaluation of a solution if further evaluations cannot allow it to escape dominance by another solution. Such approaches require the fitness function to meet some constraints regarding its statistical distribution with the racing method or its monotony with the early stopping algorithm. Besides, it does not allow to quickly discriminate between the performances of two sub-optimal solutions, which is necessary in the state-of-the-art MOEAs.

Another method consists in building an approximate model of the computationally costly fitness function, instead of computing the exact fitness value for each solution [32, 17]. Such an approach could be used to approximate the GAb of controllers on \( \Omega_{\text{eval}} \) by a so-called surrogate model, which is updated by computing the exact GAb values of only few individuals during the optimization. A similar methodology has been developed in [21] to solve the reality gap problem. It relies on a surrogate model to approximate the transfer quality from simulation to reality of controllers evolved in simulation. It makes it possible to find well-transferable controllers with few computationally costly experiments on the physical robot. Our methodology is inspired from this last work.

### 3. METHOD

The ProGAb approach deals with two goals: 1) evolving controllers with good generalisation abilities regarding several test contexts on a given task; 2) performing as few evaluations as possible on all these contexts.

Our approach is based on the three data sets methodology detailed in the section 2.1. As the performance on the training set \( \Omega_{\text{train}} \) and the GAb on the validation set \( \Omega_{\text{valid}} \) can be conflicting, we propose a multi-objective formulation of this methodology in which three objectives are optimized via a Pareto-based MOEA: (1) the performance \( F \) on the training set \( \Omega_{\text{train}} \); (2) the GAb \( G \) on the validation set \( \Omega_{\text{valid}} \); (3) the behavioral diversity objective. The addition of this third objective will be discussed later.

The GAb on the validation set can be very computationally costly to compute on every evolved solution. In order to minimize the number of evaluations on the validation set, we rely on a so-called surrogate model to approximate this second objective during the optimization process.

#### 3.1 Surrogate model of the GAb measure

Surrogate models [37, 17] are used in real engineering problems when evaluating an individual on the target system means very high computational costs or too long experiments. Then, instead of a direct evaluation on the physical system, solutions are optimized via an approximate model of it. In evolutionary algorithms, surrogate models can be used to approximate a computationally costly fitness objective. In order to build a surrogate model of the GAb function during the evolutionary optimization, a few controllers have to be assessed on \( \Omega_{\text{valid}} \) during the run to obtain the corresponding exact GAb values and update the approximate model.

To build the surrogate model, we rely on Inverse Distance Weighting interpolation [35] (IDW). While very simple, this technique has demonstrated to be competitive with more complex methods, like Kriging methods [28].

We assume that a behavioral distance between controllers \( b_{\text{dist}} \) has already been defined on the task, which compares the individuals based on their behaviors \( b(x) \) on \( \Omega_{\text{train}} \). If some controllers have already been assessed on \( \Omega_{\text{valid}} \) (at least 1) and the corresponding exact GAb values have been

<table>
<thead>
<tr>
<th>Papers</th>
<th>Reductions</th>
<th>GAb</th>
<th>( \Omega_{\text{eval}} )</th>
<th>( \Omega_{\text{test}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lehman 2010 [22]</td>
<td>x</td>
<td>1</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Ziemke 2002 [38]</td>
<td>x</td>
<td>3</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Berlanga 2002 [3]</td>
<td>x</td>
<td>6</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Doncieux 2010 [12]</td>
<td>x</td>
<td>4</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>D. Paolo 2000 [11]</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Fioravanti 2001 [14]</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Jakobi 1997 [19]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miglino 1995 [25]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jakobi 1997 [18]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kondo 1999 [20]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barate 2008 [2]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jakobi 1995 [19]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lehman 2010 [22]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miglino 1995 [25]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jakobi 1997 [18]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
computed, a surrogate model $\hat{G}$ of the generalisation ability can be interpolated by Inverse Distance Weighting from these values.

Let $\mathcal{C}$ be the set of all the possible controllers, let $\mathcal{C}_{\text{tested}}$ be the set of the controllers already assessed on $\Omega_{\text{valid}}$ and $G^*(c_i)$ the exact GAb value corresponding to each controller $c_i \in \mathcal{C}_{\text{tested}}$. The surrogate model of the GAb $\hat{G}$ is built as follows:

$$\forall c \in \mathcal{C}, \hat{G}(c) = \frac{\sum_{c_i \in \mathcal{C}_{\text{tested}}} G^*(c_i) b_{\text{dist}}(c_i, c)^{-2}}{\sum_{c_i \in \mathcal{C}_{\text{tested}}} b_{\text{dist}}(c_i, c)^{-2}}.$$  

The use of a surrogate model implies the choice of an update heuristic that selects which test experiments have to be conducted from $\Omega_{\text{valid}}$ in order to pertinently upgrade the model. We choose a relatively simple heuristic that selects, at each generation, the best controller on $\Omega_{\text{train}}$ whose behavioral distance to the controllers of $\mathcal{C}_{\text{tested}}$ is maximal. It should ensure that the surrogate model is not built on a too localized part of the behavior space.

### 3.2 Evaluation objectives

Each controller is evaluated by three objectives to be maximized:

1. the performance $F$ on $\Omega_{\text{train}}$, to find efficient controllers on the training set;
2. the corresponding approximated GAb $\hat{G}$ on $\Omega_{\text{valid}}$ computed with the surrogate model, to promote the GAb of the solutions;
3. the behavioral diversity objective.

This last objective allows to maintain behavioral diversity among the population, which efficiently enhances exploration of the controller state space [12, 26]. To quantify the diversity of a controller from the controllers already assessed on the validation set, we define a behavioral diversity value as follows. Let $\mathcal{C}_{\text{tested}}$ be the set of the controllers already assessed on $\Omega_{\text{valid}}$ and $b_{\text{dist}}$ the behavioral distance between individuals, the behavioral diversity value $\text{div}(x)$ for a given controller $x$ is:

$$\text{div}(x) = \min_{x_i \in \mathcal{C}_{\text{tested}}} b_{\text{dist}}(x, x_i).$$

This diversity value doesn’t depend on the genotype nor on the phenotype of the controller being evaluated, as the behavioral distance $b_{\text{dist}}$ is only derived from the behavior of the agent on $\Omega_{\text{train}}$. It promotes solutions that show the more different behaviors on $\Omega_{\text{train}}$ from those of the controllers already assessed on $\Omega_{\text{valid}}$ according to $b_{\text{dist}}$.

### 3.3 Algorithm

To compute the approximated GAb $\hat{G}$ at the beginning of a run, we assume that a controller $x_0$ has already been assessed on the validation set $\Omega_{\text{valid}}$. As the corresponding exact GAb $G^*(c_0)$ is known, it permits the initialization of the approximated GAb value for each controller.

The algorithm (Figure 1) iterates the following steps:

A. evaluation of the controller $x$:

1. computation of the behavior $b(x)$ and the performance $F(x)$ of $x$ on $\Omega_{\text{train}}$;
2. evaluation of the 2 other objectives in relation to the controllers of $\mathcal{C}_{\text{tested}}$ on $\Omega_{\text{valid}}$ (approximated GAb on $\Omega_{\text{valid}}$ and diversity objective);
3. application of evolutionary operators and generation of the next population.

B. depending on the update heuristic, a controller with high diversity is assessed on $\Omega_{\text{valid}}$;

C. the evolutionary operators are applied to controllers and the selection step builds the next population.

### 4. TASKS

#### 4.1 Task 1: Navigation in a T-Maze

The first task is an extension of the “roadsign problem” [38, 34]: an agent starts off at the bottom of a T-shaped maze, encounters an instruction stimulus (e.g. a light) while moving along a corridor and, when it reaches the junction, it has to turn left or right, depending on which stimulus has been encountered (Figure 2).

To make this task more cognitive in our experiment, the instruction stimulus is a combination of four stimuli (A, B, X, Y) following the same rule as in the AX-CPT working memory test [8, 31]. This task consists of a context cue (A or B), followed by a probe (X or Y) after some delay. The agent must turn to the left when the stimulus A is
followed by the stimulus X, and to the right otherwise (for AY, BX, BY). The AX-CPT task into a T-Maze has already been implemented in neuroscience experiments, to test rat’s discrimination performance [23].

Here, the agent is a simulated two-wheeled robot receiving sensory inputs from 6 infrared distance sensors and four letter sensors, one sensor for each letter A, B, X, Y, which receives 1 if the letter is presented, 0 otherwise. The robot controls its speed through two output units corresponding to its left and right motors. The agent is evaluated on each letter sequence (A followed by X, AY, BX, BY). The fitness increases by 1 if it turns to the correct side for the sequences AY, BX, BY and by 3 for the sequence AX. The maximal fitness value is 6.

To fit the experimental setup described in [23], both motors are disabled during the presentation of the letters. The whole task lasts 300 steps and takes place as follows with $t$ the number of elapsed time steps:

- $0 < t < 40$: presentation of the first letter (A/B);
- $40 \leq t < 60$: delay, all the sensors are set to 0;
- $60 \leq t < 100$: presentation of the second letter (X/Y);
- $100 \leq t \leq 300$: the robot can move and must reach the correct side of the T-maze.

### 4.2 Task 2: Ball-collecting robot

The ball-collecting task (Figure 3) has been introduced by [12]. It requires basic navigation skills—obstacle avoidance, navigation towards several goals—and the ability to change its behavior depending on the context. The goal of this task is to explore an arena, to find balls and to bring them to a basket. A fitness point is granted each time a ball is put into the basket. Before getting a point, the robot must therefore learn to search for a ball, take it and keep it, search for the basket, reach it and then, and only then leave the ball. Furthermore, the task is not completely fulfilled when the first ball is collected, as other balls have to be collected and the robot cannot carry several balls at the same time. Four balls can be collected during each experiment.

The agent is a two-wheeled robot with ten sensors: three wall distance sensors, two binary bumpers, two binary ball detection sensors, two binary basket detection sensors and one binary carry ball sensor. The effectors are left and right wheel motors and a “catch ball” actuator, whose value must be greater than 0.5 to pick up or keep a ball and lower than 0.5 to release it.

The fitness function is the number of balls in the basket at the end of the experiment. The robot has 3000 time steps to collect all the balls.

### 4.3 Control approaches

We compare ProGA$\beta$ with four control approaches:

- **Fit** only: only one fixed context is used to assess the controllers ($|\Omega_{train}| = 1$).
- **Eval** all: each individual is evaluated with all the contexts in $\Omega_{valid}$ ($\Omega_{train} = \Omega_{valid}$).
- **Fixed-Random-Initial** (FRI): $\Omega_{train}$ ($|\Omega_{train}| > 1$) is randomly generated from $\Omega_{valid}$ at the beginning of the run and remains fixed. This principle is described in [30].
- **Jakobi**: this method has been developed by [18] to provide robust controllers able to transfer well onto a physical robot. $\Omega_{train}$ is generated by randomly picking up a fixed number of contexts in $\Omega_{valid}$ at each generation. It is the same principle as in the Random-Per-Generation method (RPG) described in [30].

To be fair, all the five methods are tested with the same budget of evaluations. For the four control approaches, all the controllers of the population $P$ are evaluated on $\Omega_{train}$ during $n_g$ generations. The number of evaluations $E$ is:

$$E = size(P) \times n_g \times |\Omega_{train}|$$

In addition, for the ProGA$\beta$ approach, one controller is assessed on $\Omega_{test}$ at each generation:

$$E_{ProGA\beta} = size(P) \times n_g \times |\Omega_{train}| + n_g \times |\Omega_{valid}|$$

All the parameter values used in both tasks are described in Table 3. For ProGA$\beta$, we use the same $\Omega_{train}$ as in

---

**Figure 2:** (a) Simulated mobile robot used for the T-maze task. The robot has four additional sensors, one by letter. (b) Map employed for this task.

**Figure 3:** (a) Overview of the simulated mobile robot for the ball-collecting task. The robot is equipped with two pie-slice sensors that detect the balls and two more that detect the basket. The view range of the right sensors is depicted on the figure, the left ones are symmetric. The robot is moved by two motors, the neural network setting the speed of each of them. (b) Map employed in this set of experiments. Several initial positions and orientations of the robot are depicted with a circle and an arrow.

**Table 2:** Encoding parameters used in both tasks.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Task1</th>
<th>Task2</th>
</tr>
</thead>
<tbody>
<tr>
<td>min./max. ab. of neurons$^a$</td>
<td>3 / 20</td>
<td>10 / 30</td>
</tr>
<tr>
<td>min./max. ab. of connections$^b$</td>
<td>5 / 50</td>
<td>50 / 250</td>
</tr>
<tr>
<td>prob. to add/remove a neuron</td>
<td>0.05 / 0.05</td>
<td>0.15 / 0.05</td>
</tr>
<tr>
<td>prob. to add/remove a connection</td>
<td>0.05 / 0.05</td>
<td>0.05 / 0.05</td>
</tr>
<tr>
<td>prob. to change the weight/bias</td>
<td>0.05</td>
<td>0.15</td>
</tr>
</tbody>
</table>

$^a$Used for the initial random population.

---

**Table 3:** Encoding parameters used in both tasks.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Task1</th>
<th>Task2</th>
</tr>
</thead>
<tbody>
<tr>
<td>prob. to change the weight/bias</td>
<td>0.05</td>
<td>0.15</td>
</tr>
<tr>
<td>prob. to add/remove a connection</td>
<td>0.05 / 0.05</td>
<td>0.05 / 0.05</td>
</tr>
<tr>
<td>prob. to add/remove a neuron</td>
<td>0.05 / 0.05</td>
<td>0.15 / 0.05</td>
</tr>
<tr>
<td>min./max. nb. of connections</td>
<td>5 / 50</td>
<td>50 / 250</td>
</tr>
<tr>
<td>min./max. nb. of neurons</td>
<td>3 / 20</td>
<td>10 / 30</td>
</tr>
<tr>
<td>min./max. nb. of neurons</td>
<td>3 / 20</td>
<td>10 / 30</td>
</tr>
<tr>
<td>min./max. nb. of connections</td>
<td>5 / 50</td>
<td>50 / 250</td>
</tr>
<tr>
<td>prob. to change the weight/bias</td>
<td>0.05</td>
<td>0.15</td>
</tr>
<tr>
<td>prob. to add/remove a neuron</td>
<td>0.05 / 0.05</td>
<td>0.15 / 0.05</td>
</tr>
<tr>
<td>prob. to add/remove a connection</td>
<td>0.05 / 0.05</td>
<td>0.05 / 0.05</td>
</tr>
<tr>
<td>prob. to change the weight/bias</td>
<td>0.05</td>
<td>0.15</td>
</tr>
</tbody>
</table>
Fit\textsubscript{Only}. A budget of 3,800,000 evaluations is arbitrarily fixed for the T-maze task, which corresponds to 10,000 generations for ProGAb. A budget of 1,200,000 evaluations is arbitrarily fixed for the ball-collating task, i.e. 5,500 generations for ProGAb.

The parameters for each context \( \omega \) in \( \Omega_{\text{train}}, \Omega_{\text{valid}}, \Omega_{\text{test}} \) for both tasks are: 1) the map size; 2) the initial orientation of the robot; 3) the initial position of the robot. We add another parameter in the ball-collating task: the initial positions of the balls. There are 6 ball positions for 4 balls which means there are 15 different contexts (\( C^6_4 \)) for a given map size and a given initial position. These parameters are described in Table 4.

All the five approaches have been implemented using the efficient state-of-the-art MOEA NSGA-II \([10]\). Both tasks are simulated in a simple 2D simulator without friction or slippage.

Diversity objective.

To efficiently explore the controller state space, we add a behavioral diversity objective to the fitness function in a Pareto-based multi-objective optimization for all the four control approaches: Fit\textsubscript{Only}, Eval\textsubscript{all}, FRI and Jakobi. According to the conclusions of \([9]\), the behavioral diversity objective to maximize \( o_{bd}(x) \) is the average behavioral distance to the rest of the population \( P \). Let \( b_{\text{dist}} \) be the behavioral distance used to compare the individuals and \( F \) the fitness function, the controllers are optimized via the following multi-objective scheme:

\[
\text{maximize} \quad \left\{ \frac{F(x)}{o_{bd}(x) + \frac{1}{|\text{set}(x)|} \sum_{y \in \text{set}(x)} b_{\text{dist}}(x,y)} \right\}
\]

Neural networks.

Neural networks are evolved structurally and parametrically to control a robot with a simple direct encoding \([26]\), inspired by NEAT \([36]\). Classic sigmoid neurons are used with an output in \([-5, 5]\) and a bias in \([-5, 5]\).

In this encoding, a neural network is described as a directed graph and five mutation operators are implemented: adding or deleting a connection; adding or deleting a neuron; changing a weight or a bias using polynomial mutation \([10]\). Cross-over is not employed. All the parameter values used in both tasks are described in Table 2. Further details on the neural encoding can be found in \([56]\).

Behavioral distance.

We use Hamming distance as behavioral distance \( b_{\text{dist}} \). This generic distance has been successfully used in \([12]\), where the Hamming-based approaches generated the most efficient controllers. The Hamming distance counts the number of bits that differ between two binary sequences. In both tasks, we use the binarised sequence of all the sensor and effector values.

\begin{table}[h]
\begin{center}
\begin{tabular}{|c|c|c|c|c|}
\hline
\textbf{Method} & \textbf{size(F)} & \textbf{Task 1:} \( \Omega_{\text{valid}} = 180 \) & \textbf{Task 2:} \( \Omega_{\text{valid}} = 30 \) \\
\hline
\text{Fit\textsubscript{Only}} & 200 & 10000 & 1 & 6000 & 1 \\
\text{FRI} & - & 1980 & 10 & 1500 & 4 \\
\text{Jakobi} & - & 1980 & 10 & 1500 & 4 \\
\text{Eval\textsubscript{all}} & - & 110 & 180 & 200 & 30 \\
\text{ProGAb} & - & 10000 & 1 & 5500 & 1 \\
\hline
\end{tabular}
\end{center}
\end{table}

\begin{table}[h]
\begin{center}
\begin{tabular}{|c|c|c|}
\hline
\textbf{Method} & \textbf{size(F)} & \textbf{map size} \\
\hline
\text{Fit\textsubscript{Only}} & 200 & 600 to 650 \\
\text{Eval\textsubscript{all}} & - & 550 to 650 \\
\text{FRI} & - & 30 to 300 \\
\text{Jakobi} & - & -20 to 20 \\
\text{ProGAb} & - & -20 to 20 \\
\hline
\end{tabular}
\end{center}
\end{table}

5. RESULTS

The following results are investigated: 1) the evolution of the GAb on \( \Omega_{\text{valid}} \) for the best controllers obtained with each method (Figures 5 and 7); 2) the GAb on \( \Omega_{\text{valid}} \) and \( \Omega_{\text{test}} \) of the best-of-run individual (Figures 4 and 6). All the comparisons between the methods are made with Wilcoxon rank-sum tests.

5.1 Task 1: navigation in a T-maze

For the ProGAb and Fit\textsubscript{Only} approaches, the maximum fitness on \( \Omega_{\text{train}} \), is reached quickly, with respective median values of 5500 and 71000 evaluations (results not pictured). The best controllers obtained with \( \text{Fit\textsubscript{Only}} \) achieve a median of successful contexts on \( \Omega_{\text{valid}} \) of 3 out of 180, while a median of 170 successful contexts is observed for ProGAb. Such results highlight that reaching the maximal performance on only one context is not sufficient to obtain general controllers, as it does not give any information on the GAb on \( \Omega_{\text{valid}} \).

Performing a lot of evaluations is not necessary relevant: there is no significant difference between the GAb on \( \Omega_{\text{valid}} \) after 1.9 and 3.8 millions of evaluations for all the methods (p-values > 0.35).

Increasing the number of contexts in \( \Omega_{\text{train}} \) improves the GAb: Jakobi and FRI perform better than \( \text{Fit\textsubscript{Only}} \) (p-values < 10\(^{-3}\)). But assessing on a huge number of contexts requires a lot more evaluations: with a budget of 3.8 millions evaluations, Eval\textsubscript{all} performs clearly worse than FRI or Jakobi (p-values < 2·10\(^{-5}\)).

Randomly generating \( \Omega_{\text{train}} \) from \( \Omega_{\text{valid}} \) at each generation does not improve significantly the GAb on \( \Omega_{\text{valid}} \); the p-value obtained by comparing FRI and Jakobi is 0.15 after 3.8 millions of evaluations, with respective median values of 80 and 153.

\begin{table}[h]
\begin{center}
\begin{tabular}{|c|c|c|c|}
\hline
\textbf{Method} & \textbf{Number of successful runs (out of 20)} & \textbf{Evaluated before convergence} & \textbf{Mean} \\
\hline
\text{Fit\textsubscript{Only}} & 6 & 2.59 & 2.59 & 0 \\
\text{Eval\textsubscript{all}} & 1 & 1.75 & 1.86 & 0.95 \\
\text{FRI} & 6 & 1.21 & 1.64 & 1.12 \\
\text{Jakobi} & 12 & 0.42 & 0.84 & 1.93 \\
\text{ProGAb} & 14 & & & \\
\hline
\end{tabular}
\end{center}
\end{table}
The ProGAb approach finds better general controllers than the other methods with a median of successful contexts on $\Omega_{valid}$ of 170 out of 180 ($p$-values $< 4 \cdot 10^{-2}$). Moreover, these controllers are found quicker than with the other methods. If we look at the runs where the individual’s GAb is greater than 170 out of 180 contexts (Table 5), we see that ProGAb needs about two to three times fewer evaluations than Jakobi to converge, with respective median values of 0.43 and 1.20 millions of evaluations ($p$-value $= 3 \cdot 10^{-2}$). At last, there is no significant difference between the GAb obtained with the ProGAb approach on $\Omega_{valid}$ and on $\Omega_{test}$ ($p$-value $= 0.45$), with respective median values of 170 and 150 successful contexts out of 180. It emphasizes the effective GAb of the controllers obtained with ProGAb.

5.2 Task 2: ball-collecting task

The Figure 6 shows the average number of balls collected by the best controllers on 20 runs obtained with each method on the three sets. The FiF-only approach does not always reach the maximal performance on $\Omega_{train}$ (the four balls are collected in only 13 runs out of 20). It is consistent with previous experiments on this setup [12] and reflects the complexity of the ball-collecting task.

The ProGAb approach finds controllers with significantly better GAb than the four other approaches, on $\Omega_{valid}$ (p-value $< 10^{-2}$) as well as on $\Omega_{test}$ (p-value $< 10^{-2}$). Moreover, according to Figure 6, the GAb on $\Omega_{valid}$ (median = 2.63 balls) and on $\Omega_{test}$ (median = 2.23 balls) are not significantly different (p-value = 0.11) for this approach. Contrarily, the GAb significantly decreases for Jakobi’s method, as the median drops from 2.13 on $\Omega_{valid}$ to 0.74 on $\Omega_{test}$ (p-value $< 5 \cdot 10^{-2}$).

Several conclusions can be drawn from the results obtained on both tasks:

- evaluating with a single context or a few fixed contexts does not lead to general controllers;
- changing the training set from a generation to another can lead to controllers with a good GAb if the number of evaluations is sufficiently high;
- the ProGAb approach obtains significantly better general controllers, while requiring fewer evaluations.

6. CONCLUSION

This paper proposes the ProGAb approach which promotes the generalisation abilities (GAb) of controllers. It is based on the three data sets methodology used in supervised machine learning and relies on a surrogate model for approximating the GAb of the controllers in order to cope with computational cost issues.

In the first experiment where a robot must turn to the correct side of a T-maze according to a previously encountered stimulus, ProGAb converges quicker than the other methods. With the more complex ball-collecting task, where a robot must explore an arena to find balls and take them to a basket, ProGAb is both quicker to converge and provides controllers with better generalisation abilities.

Based on these successful results, our approach appears to be a simple and relevant method to efficiently evolve controllers with good generalisation abilities.

7. ACKNOWLEDGEMENT

This project was funded by the ANR-09-EMER-005-01 project.
8. REFERENCES