From Evolutionary Robotics to Developmental Robotics

ISIR, UPMC-CNRS
stephane.doncieux@isir.upmc.fr

LIRIS, 16/12/2014

ISIR : Institut of Intelligent Systems and Robotics

- created in 2007
- 52 researchers
- 56 phd students
- 35 post-docs

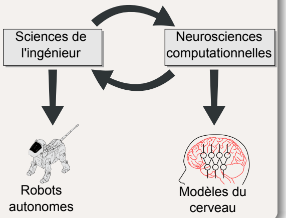
Director : R. Chatila
4 research teams :

- AGATHE : human movement assistance with a robotic device
- INTERACTION : haptic, human-robot interaction, interacting with the micro-world
- SYROCO : mobile robots in complex environments
- AMAC : Architectures and models of adaptation and cognition

AMAC : Architectures and models of adaptation and cognition

Created in 2011. Resp : S. Doncieux
Staff : 12 researchers, 15 phd students, 3 post-docs
Research questions :

- What learning or adaptation processes to acquire new skills and adapt them to the context ?
- What cognitive architecture to manage the different learning processes and exploit at best the learned skills ?



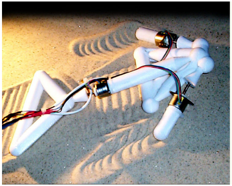
AMAC research topics

Computational neuroscience axis

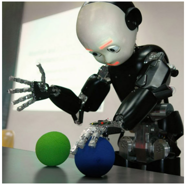
- posture and motion control
- action selection
- spatial navigation
- reward based learning
- modeling : from neuron properties to global architectures

Engineering science axis

- Adaptation mechanisms for complex robotics systems :
  - reinforcement learning
  - model learning
  - multi-objective evolutionary robotics
- From low-level control to discrete actions
- Cognitive architecture integrating learning, perception, interpretation, decision and action
- Selection pressures for evolutionary robotics







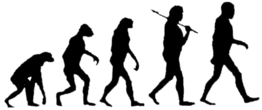
Motivation

- Building robots with *embodied intelligence* [Pfeifer 2007]

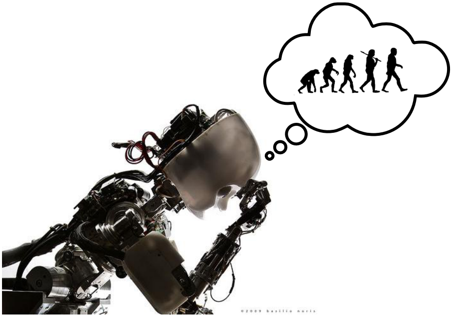
Positioning

- When you know how to exhibit the behavior of interest :
  - Behavior based robotics [Brooks 1986, Mataric and Michaud 2008]
- When you know what to do :
  - Learning by imitation [Schaal et al. 2003]
  - Supervised learning [Pomerleau 1991]
  - Inverse RL [Ng and Russel 2000]
- When you can evaluate what is done :
  - Model-based reinforcement learning [Sutton and Barto 1998]
  - **Policy search (evolutionary) RL** [Nolfi and Floreano 2000]

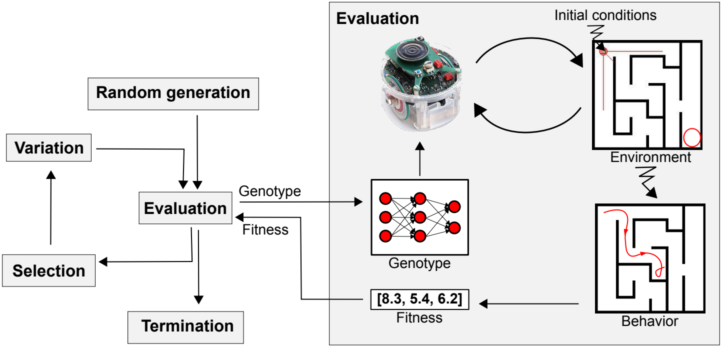
Overview



1. ISIR contributions to Evolutionary Robotics : Selective Pressures
2. Evolutionary Robotics and Development



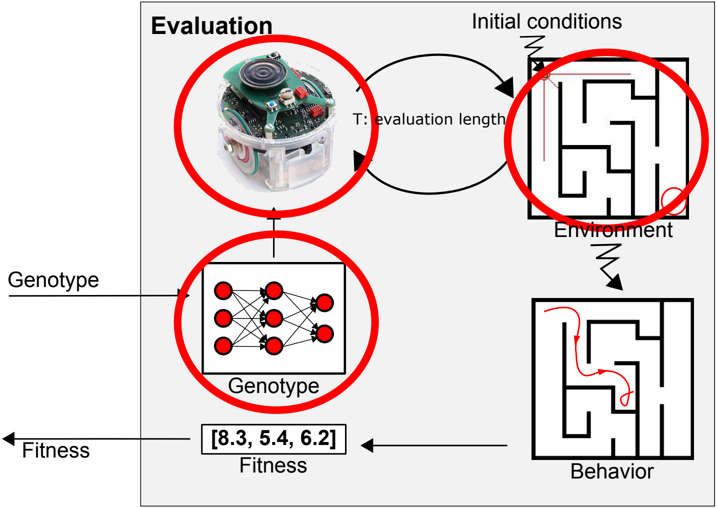
Evolutionary Algorithms & Robotics



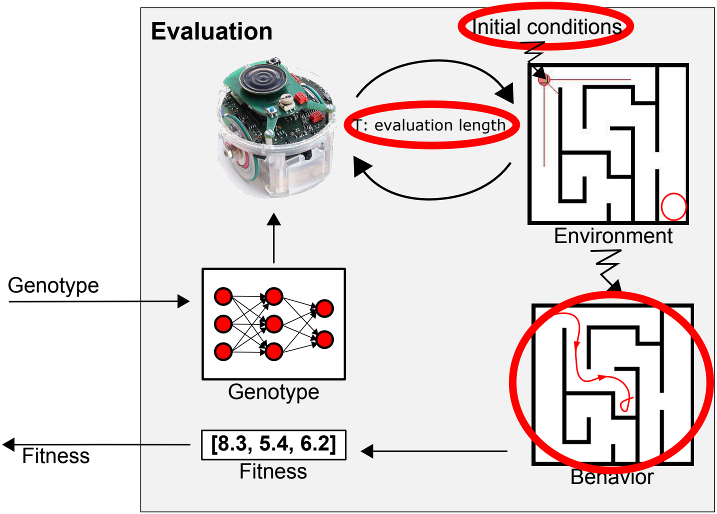
Eiben, A.E., Smith, J.E.(2007)
Introduction to Evolutionary Computing
ISBN : 978-3-540-40184-1. Springer

- stochastic, population based algorithms
- converge to an approximation of the optimal solution
- can do more than parameter optimization
- $f(.)$ need not be differentiable, nor continuous
- $f(.)$ may even not be known analytically, but measured through a specific device
- $f(.)$ may be noisy, multi-modal
- $f(.)$ may be in $\mathbb{R}^n$

Evolutionary Algorithms & Robotics

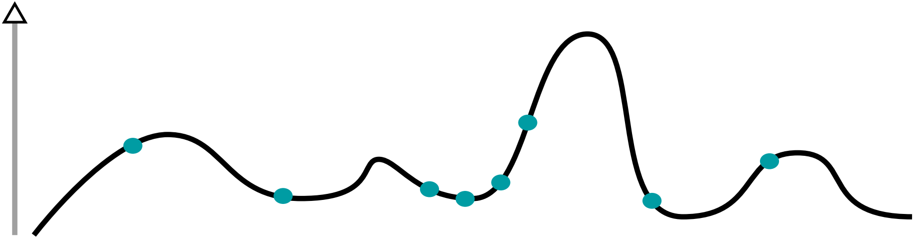


Evolutionary Algorithms & Robotics

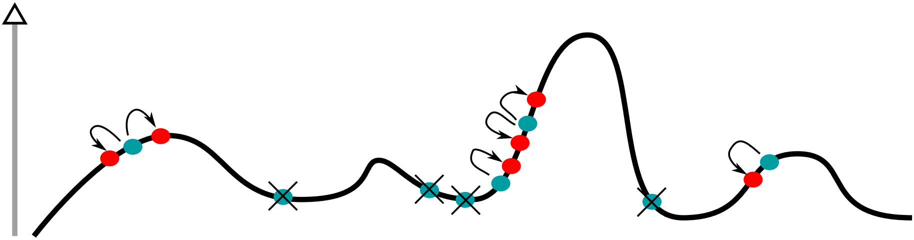


There is room for the definition of generic selective pressures !

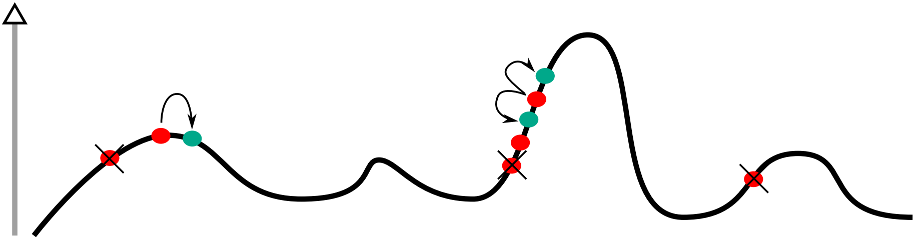
Selective pressures



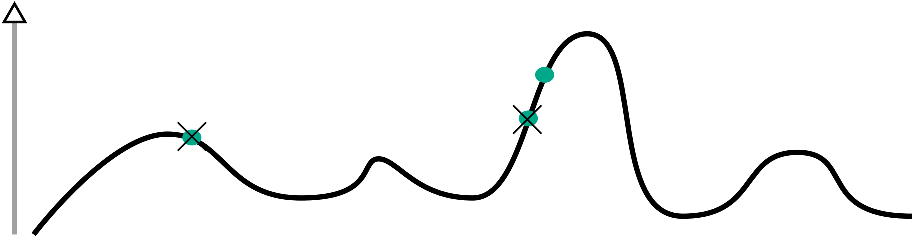
Selective pressures



Selective pressures



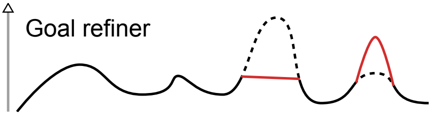
Selective pressures

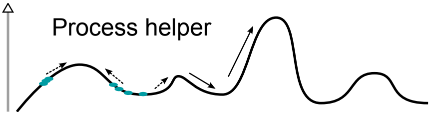


The fitness has 2 different roles :

1. it defines the goal
2. it guides the search

Classification of selective pressures





Goal refiners

A goal refiner aims at changing the optimum(s) of the fitness function by adding new requirements.

Process helpers

A process helper intends to increase the efficiency of the search process without changing the optimum(s) of the fitness function.

Task specific

Task specific goal refiners/process helpers incorporate knowledge on how to solve the task.

Task agnostic

Task agnostic goal refiners/process helpers do not exploit knowledge about how to solve the task.

Doncieux, S., and Mouret, J.-B. (2014).
Beyond Black-Box Optimization : a Review of Selective Pressures for Evolutionary Robotics.
Evolutionary Intelligence. doi :10.1007/s12065-014-0110-x

ER@ISIR : Learning as a multi-objective optimization

Approach : neuroevolution

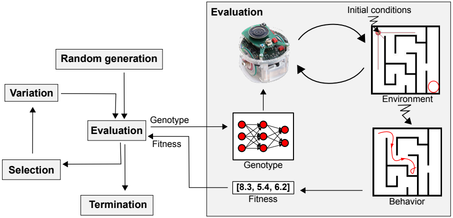
Direct policy search : learning in policy space
$\rightarrow$ **continuous** space, perceptions and actions
$\rightarrow$ control architecture **X** : **neural networks**.
$\rightarrow$ stochastic optimization algorithms : **evolutionary algorithms**.

Find a robot controller **X** (policy)

maximizing : $\mathbf{f}(\mathbf{X}) = \begin{Bmatrix} f_1(\mathbf{X}) \\ f_2(\mathbf{X}) \\ \vdots \\ f_n(\mathbf{X}) \end{Bmatrix}$

$\rightarrow$ **process helpers as new objectives**

$\rightarrow$ **goal refiner as new objectives**



Doncieux, S., and Mouret, J.-B. (2014).
Beyond Black-Box Optimization : a Review of Selective Pressures for Evolutionary Robotics.
Evolutionary Intelligence. doi :10.1007/s12065-014-0110-x

Process helper : Premature convergence

Intensification vs diversification in evolutionary algorithms

- Exploration : stochastic search operators & population ;
- exploitation : fitness function.

Hypothesis

The bottleneck of ER may be due to an exploration problem.

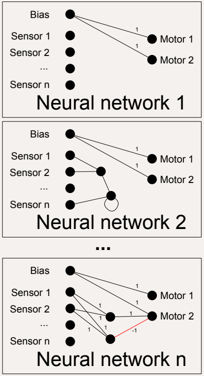
Intensification vs diversification in EA

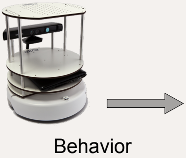
How to keep a diverse population ?

$\rightarrow$ by penalizing similar individuals on the basis of their genotype or phenotype :

- fitness sharing [Goldberg and Richardson 1987]
- objective on diversity in a multi-objective scheme [Abbas and Deb 2003, de Jong et al. 2001]
- niches [Sareni and Krähenbühl 1998]

Process helper : Premature convergence





n neural networks $\rightarrow$ 1 behavior

Why not promoting diversity in the space of behaviors ?

[Lehman and Stanley 2008] [Trujillo et al. 2008] [Mouret and Doncieux 2009] [Gomez 2009]

Process helper, task agnostic

Process helper : Premature convergence

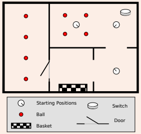
How to describe and compare behaviors ?

- adhoc descriptions :
  - final position [Lehman and Stanley 2008]
  - environment state [Mouret and Doncieux 2009]
- generic descriptions :
  - robot trajectory [Trujillo et al 2008]
  - hamming distance [Doncieux and Mouret 2010]
  - entropy [Delarboulas et al. 2011]

ISIR approach, multi-objectivization : Behavioral diversity

Our proposition :

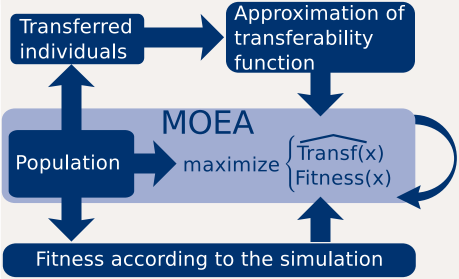
Find $g_{\mathbf{X}}(.)$ maximizing $\begin{cases} f(g_{\mathbf{X}}(.)) \\ \frac{1}{N}\sum_{j=0}^{j=N} d(g_{\mathbf{X}}(.), g_{\mathbf{Y}}(.)) \end{cases}$

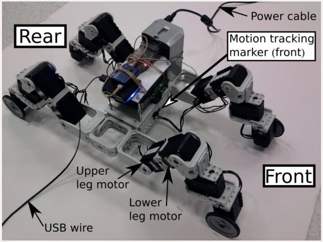


with $d(\mathbf{X}, \mathbf{Y})$ behavioral distance between $\mathbf{X}$ and $\mathbf{Y}$

Mouret, J.-B. and Doncieux,S. (2012)
Encouraging Behavioral Diversity in Evolutionary Robotics : an Empirical Study
Evolutionary Computation. Vol 20 No 1 Pages 91-133.

Goal refiner : Reality gap





ISIR approach : multi-objectivization

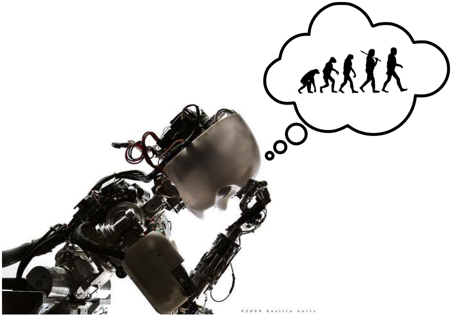
$$\text{Find } g_{\mathbf{x}}(.) \text{ maximizing } \begin{cases} f(g_{\mathbf{x}}(.)) \\ \widehat{Transf}(g_{\mathbf{x}}(.)) \end{cases}$$

Koos, S. and Mouret, J.-B. and Doncieux S. (2013).
The Transferability Approach : Crossing the Reality Gap in Evolutionary Robotics.
IEEE Transaction on Evolutionary Computation. Vol 17 No 1 Pages 122 – 145.

Goal refiner, task agnostic

From Evolutionary Robotics to Developmental Robotics

From Evolutionary Robotics to Developmental Robotics

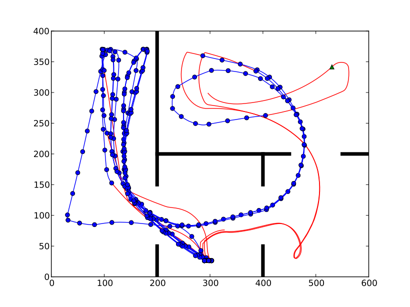


Darwinian selection process inside the brain ?

Hypothesis explored by several authors :

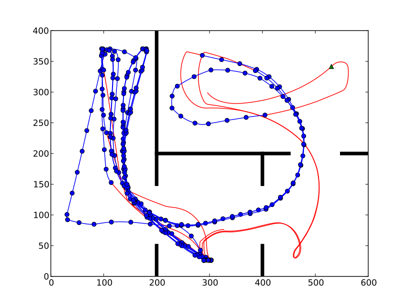
- Changeux, J. P., et al. (1984). Learning by selection, in The Biology of Learning, Springer, 115–133.
- Edelman, G. M. (1987). Neural Darwinism. The Theory of Neuronal Group Selection. New York : Basic Books.
- Fernando, C. et al. (2012). Selectionist and evolutionary approaches to brain function : a critical appraisal Frontiers in Computational Neuroscience, 6(April), 1–28

Back to collectball experiment

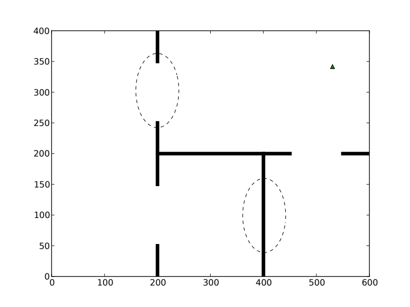


On the environment used for evaluation

Back to collectball experiment

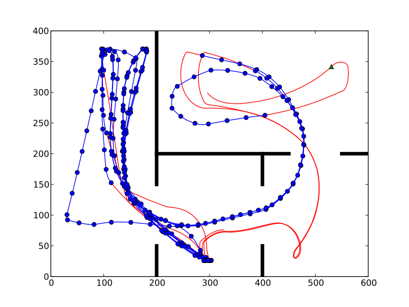


On the environment used for evaluation

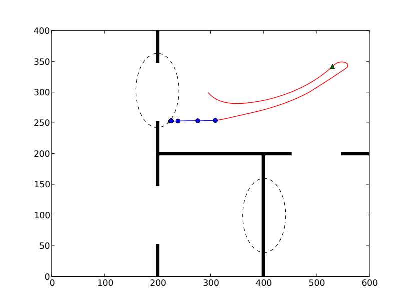


On a new environment (modifications of walls only)

Back to collectball experiment



On the environment used for evaluation



On a new environment (modifications of walls only)

→ The robot doesn't understand what it is doing...

Towards lifelong learning

New situation with current ER approaches

- controller robust enough : nothing to do !
  - Promoting generalization ability [Pinville et al. 2011]
  - Promoting reactivity [Lehman et al. 2013]
- current controller fails :
  - What to do ? New learning episode from scratch ?

Questions

- What knowledge to transfer between learning episodes ?
- How to transfer knowledge ?

Towards lifelong learning

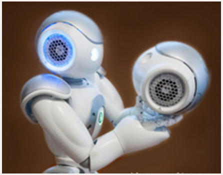
New situation with current ER approaches

- controller robust enough : nothing to do !
  - Promoting generalization ability [Pinville et al. 2011]
  - Promoting reactivity [Lehman et al. 2013]
- current controller fails :
  - What to do ? New learning episode from scratch ?

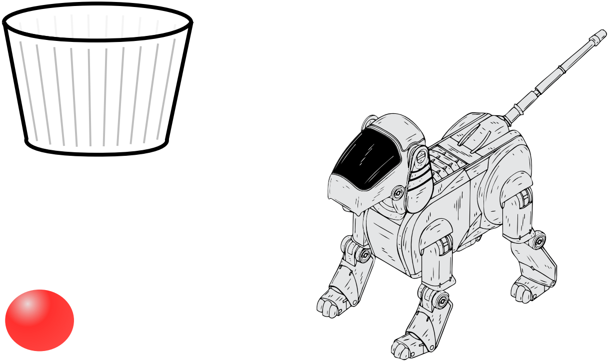
Questions

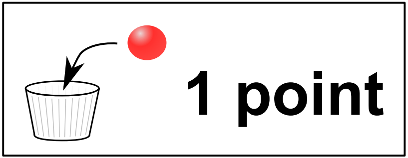
- What knowledge to transfer between learning episodes ?
- How to transfer knowledge ?

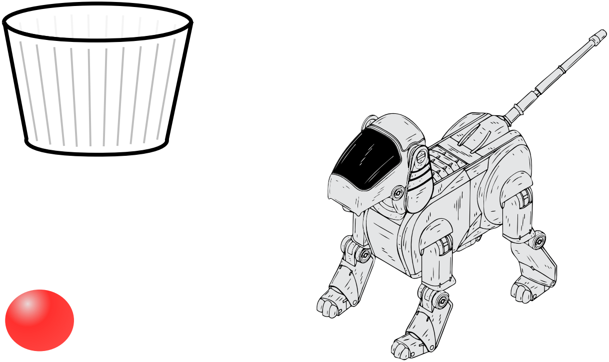
What makes sense for a robot ?

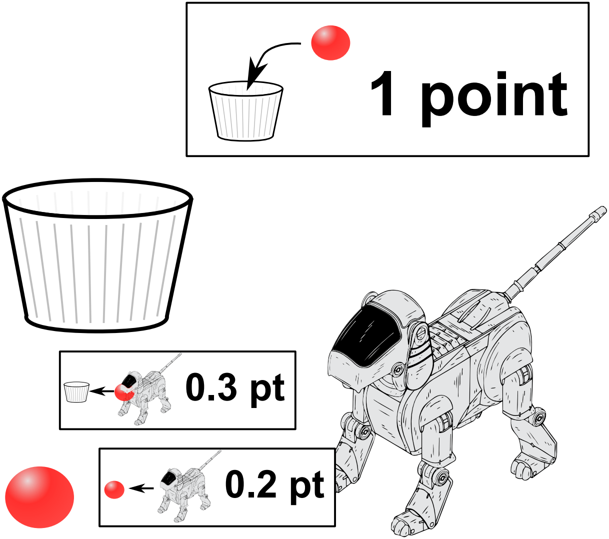


- with regards to a particular environment
- given a particular morphology
- given some drives/motivations/tasks




1 point




1 point
0.3 pt
0.2 pt

Proposed approach

Hypothesis

- numerous attempts to satisfy the drive/motivation/task, some being successful, and some failing ;

→ **meaningful concepts (actions, perceptions) are those that are specific to successful attempts.**

Methodological issue

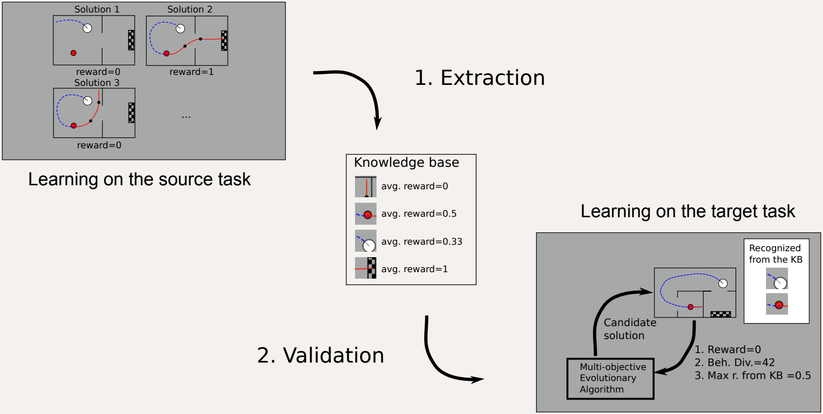
- How to validate that the knowledge acquired in a particular context is *meaningful* ?

→ by proving that this knowledge is useful in a new and similar context :

transfer learning problem [Thrun 1996]

What makes sense ?

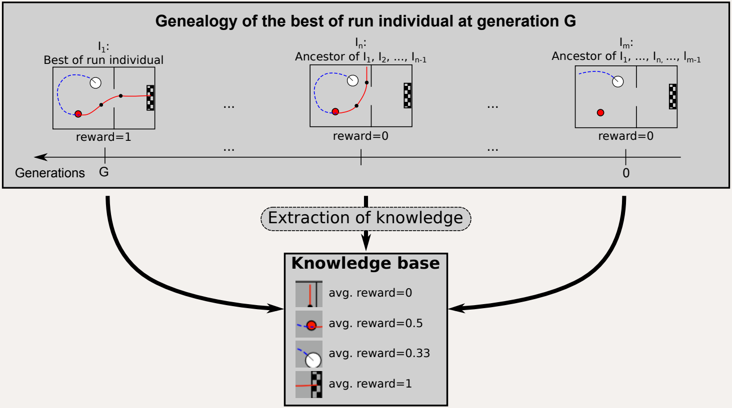
Overview



Doncieux, S. (2013).
Transfer Learning for Direct Policy Search : A Reward Shaping Approach,
In Proceedings of the IEEE ICDL-EpiRob conference.

What makes sense ?

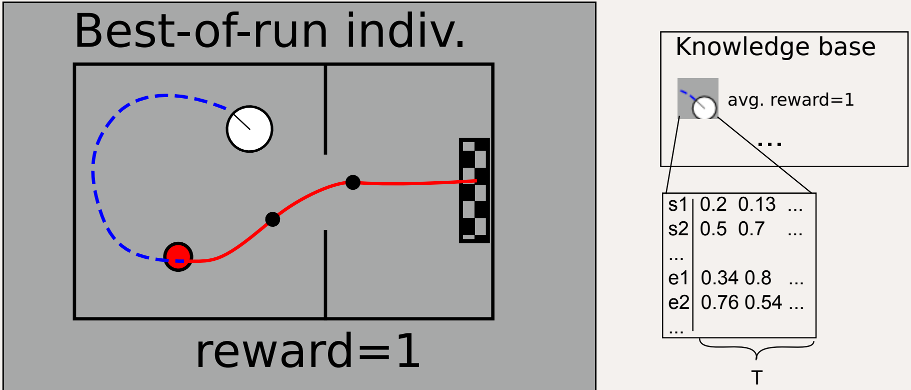
Knowledge Extraction from Best-of-run Individual



Doncieux, S. (2014).
Knowledge Extraction from Learning Traces in Continuous Domains.
AAAI 2014 fall Symposium "Knowledge, Skill, and Behavior Transfer in Autonomous Robots".

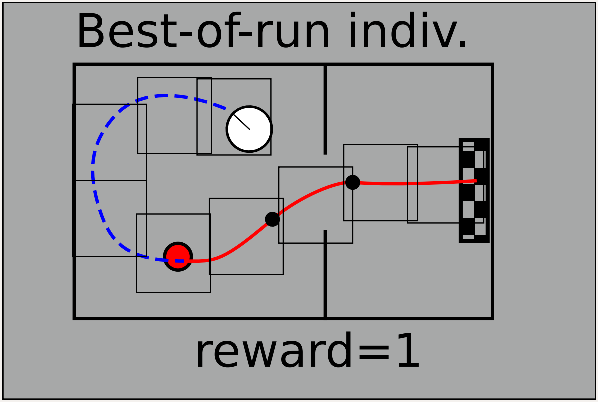
What makes sense ?

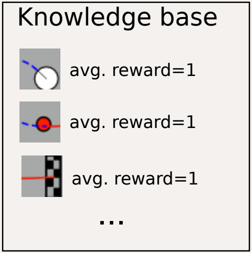
1. Extraction : learning on the source task



What makes sense ?

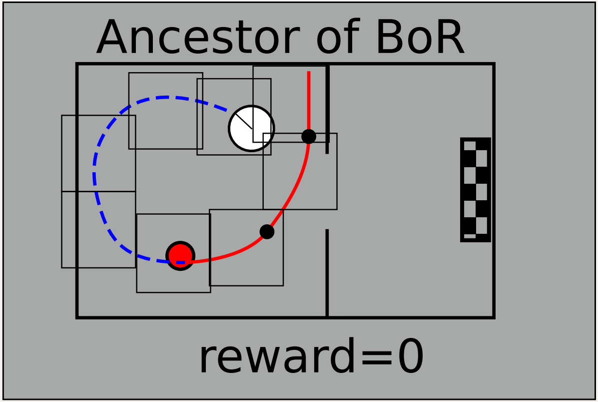
1. Extraction : learning on the source task

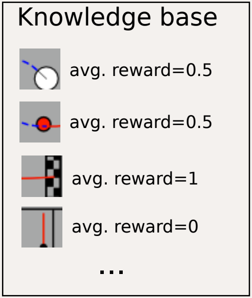




What makes sense ?

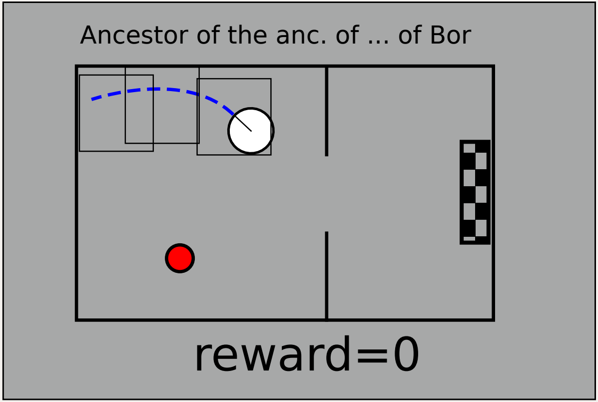
1. Extraction : learning on the source task

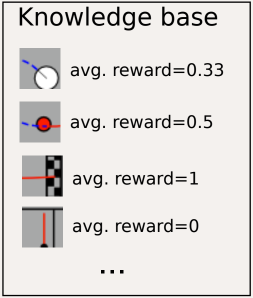




What makes sense ?

1. Extraction : learning on the source task





What makes sense ?

2. Validation : transfer learning on the target task

Knowledge base

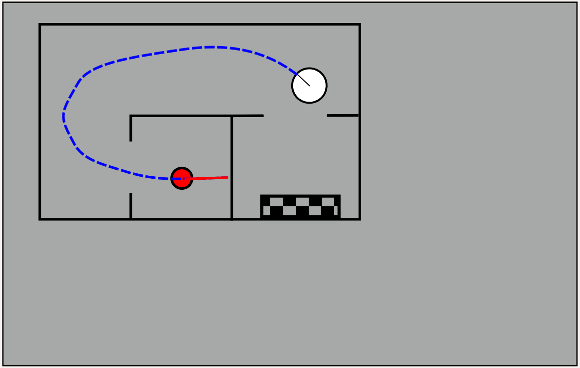
avg. reward=0.33

avg. reward=0.5

avg. reward=1

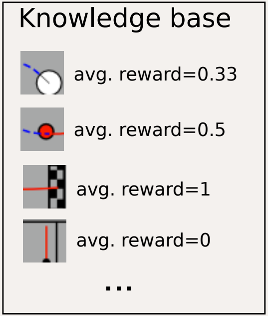
avg. reward=0

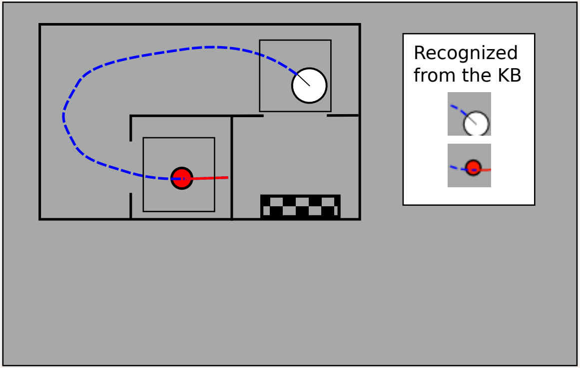
...



What makes sense ?

2. Validation : transfer learning on the target task





What makes sense ?

2. Validation : transfer learning on the target task

Knowledge base

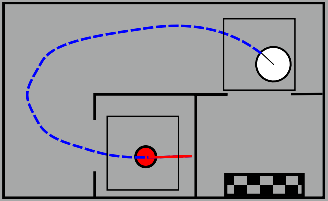
 avg. reward=0.33

 avg. reward=0.5

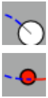
 avg. reward=1

 avg. reward=0

...



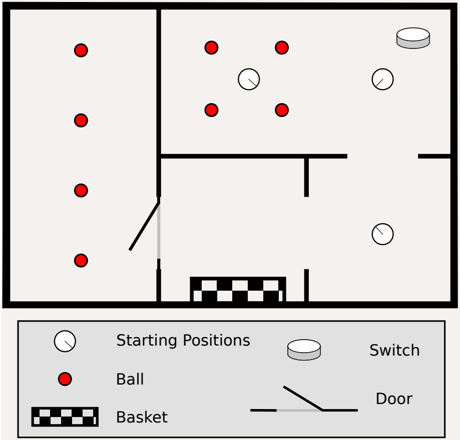
Recognized from the KB



Objectives:

1. Reward=0
2. Beh. Div.=42
3. **Max r. from KB =0.5**

Experimental setup : source task (1)

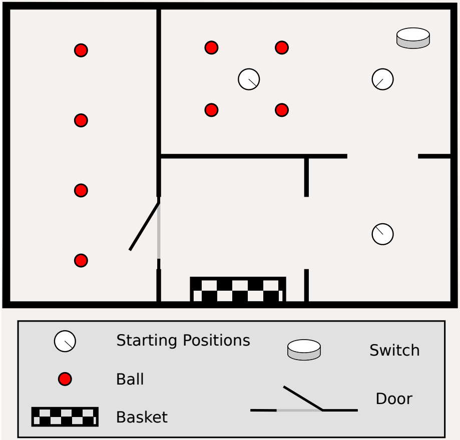


Setup :

- 12 inputs :
  - 3 laser range finders
  - 2 ball sensors
  - 2 basket sensors
  - 2 switch sensors
  - 2 bumpers
  - 1 carrying ball sensor
- 3 outputs
  - 2 motors
  - 1 collect ball/launch ball effector

Knowledge base : KD-tree, as implemented in the FLANN library

Experimental setup : source task (2)

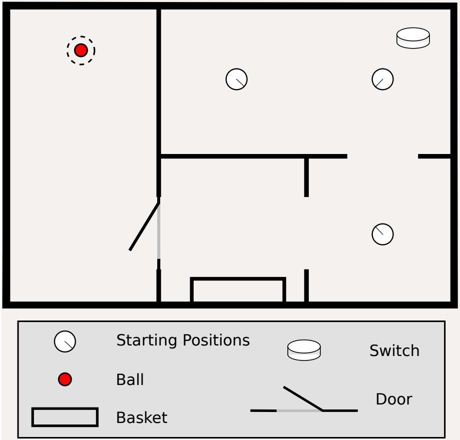


Setup :

- Multi-objective fitness :
  1. number of collected balls
  2. dynamic beh. div. [1]
- EA : NSGA-II, pop size : 200, nb gen : 4000
- direct neural network encoding :
  - up to 30 neurons
  - 50 to 250 connections
  - activation function : $y_i = \varphi\left(\sum_j w_{ij} x_j\right)$ where $\varphi(x) = \frac{1}{1+\exp(b-kx)}$

[1] Doncieux, S. and Mouret, J.B. (2013). .
Behavioral Diversity with Multiple Behavioral Distances.
Proc. of IEEE Congress on Evolutionary Computation, 2013 (CEC 2013).

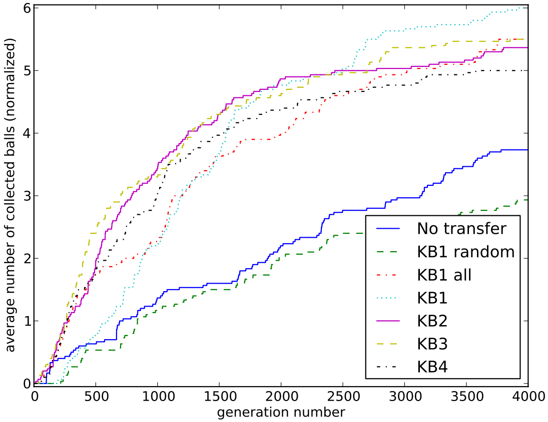
Experimental setup : target task (1)



Setup :

- Same inputs/outputs than for the source task.
- EA : NSGA-II, pop size : 200, nb gen : 4000
- direct neural network encoding :
  - up to 30 neurons
  - 50 to 250 connections
  - activation function : $y_i = \varphi\left(\sum_j w_{ij} x_j\right)$ where $\varphi(x) = \frac{1}{1+\exp(b-kx)}$

Results



Average fitness value (30 runs)

Setup with transfer

KB1, ..., KB4, objectives :

1. Goal
2. Diversity
3. Max reward from KB

→ KB generated from diff. exp.

KB1all, objectives :

1. Goal
2. Diversity
3. Max reward from KB

→ KB generated from all indiv.

Control experiments

no transfer, objectives :

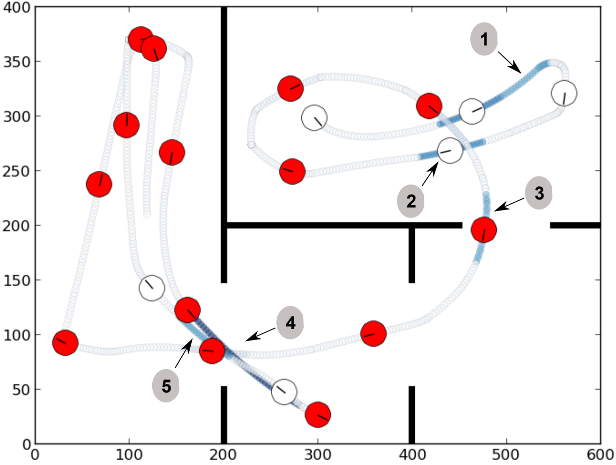
1. Goal
2. Diversity

KB1 random, objectives :

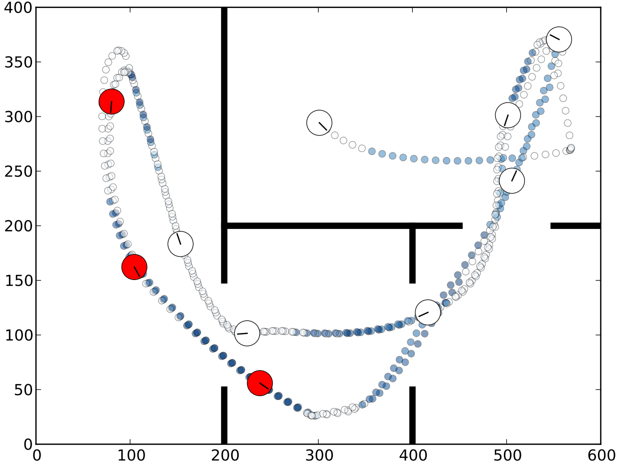
1. Goal
2. Diversity
3. Max reward from KB

→ KB = KB1 with random values

Source task : What makes sense for the robot ?



Target task : What makes sense for the robot ?



Conclusion

Take home messages

- learning as a multi-objective optimization problem
- learning in continuous domains thanks to neuroevolution
- **knowledge can emerge from the analysis of successes and failures**

Perspectives : developmental robotics

- Building a repertoire of primitive actions/perceptions for use with other and faster learning algorithms
- Use extracted values to focus attention to specific features

→ **H2020 FET European project DREAM (2015-2018)**

ENSTA ParisTech ARMINES Queen Mary University of London vrije Universiteit amsterdam UPMC SORBONNE UNIVERSITÉS CNRS UNIVERSIDADE DA CORUÑA

Papers

Doncieux, S. and Mouret J.-B. (2014).
Beyond Black-Box Optimization : a Review of Selective Pressures for Evolutionary Robotics.
Evolutionary Intelligence Journal.

Doncieux, S. (2014).
Knowledge Extraction from Learning Traces in Continuous Domains.
AAAI 2014 fall Symposium ”Knowledge, Skill, and Behavior Transfer in Autonomous Robots”.

Doncieux, S. (2013).
Transfer Learning for Direct Policy Search : A Reward Shaping Approach.
Proceedings of ICDL-EpiRob conference. Pages 1-6.

Doncieux, S., Bredeche, N., Mouret, J.-B. and Eiben, A.E. (submitted).
Evolutionary Robotics : what, why and where to.

Source code

Sferesv2 software framework : `https://github.com/jbmouret/sferes2`

Source code of the experiments : `http://www.isir.fr/evorob_db`

Thank you for your attention !

Email : `stephane.doncieux@upmc.fr`