Open-ended (re-)search in developmental and evolutionary systems

Pierre-Yves Oudeyer

INRIA FLOWERS

http://www.pyoudeyer.com
Developmental/epigenetic systems

• Developmental/epigenetic systems are systems which, given innate materials/structures/properties/mechanisms/capabilities, grow as a result of the dynamical interactions between the components of this innate equipment among themselves and with their environment;
  ➔ Developmental processes = these growth processes

• In biological organisms, these interactions take place simultaneously at many organizational levels: 
  e.g. at the level of individual cells through for e.g. differentiation processes resulting from dynamical patterns of gene expression that depend on the chemical environment within and outside the cell;
  or at the level of cell assemblies which dynamically interact among themselves and with the physical environment to generate organs (e.g. through interaction with gravity), or for e.g. to learn immune responses, or to maturate for e.g. neural systems:
  or at the behavioral level of the whole organisms which neural and learning system interact with the physical and social environment, to generate for e.g. locomotion or language;

• Not only biological systems can be developmental, e.g. galaxies or universes;
• Developmental/epigenetic robotics ➔ study of bio-inspired growth processes;
Developmental processes involve complex (physico-chemico) dynamical systems, both in embryogenesis and epigenesis

⇒ Self-organization phenomena

⇒ Constrains the space of forms and structures: not all forms are equally easy to generate from physical matter/are possible ⇒ Constrains evolutionary search

⇒ Evo-devo approach both for biological understanding AND engineering efficiency


One non-standard bridges between research on evolutionary and developmental systems: the role of open-ended search mechanisms
4 types of \textit{(re-)}search in (computational/mathematical) evolutionary and developmental systems

\textbf{Evolutionary systems:}
1) Task-specific optimization problems for engineering purposes;
2) Modeling work for understanding better the origins of structures and complexity in phylogeny/biological evolution;

\textbf{Developmental systems:}
1) Task-specific learning problems for engineering purposes;
2) Modeling work for understanding better the formation of structures in sensorimotor/cognitive/social development;

- These four types of research can be formulated in a unified formal framework;
- Yet, very different goals, and underlying technical differences that lie mainly in the kind of search process they are associated with;
- In spite of strong differences in objectives, recent results show how strong synergies can relate them;
A common underlying formulation

• A (discrete or continuous) state space $S$ (e.g. sensori state and memory of a robot)
• A (discrete or continuous) action space $A$ (e.g. motor commands of a robot)
• A transition function $W : S(t) \times A(t) \rightarrow S(t+1)$
• A parameterized action policy $\pi_\theta : S \rightarrow A$

• A reward/value/fitness function
  $R : S \rightarrow \mathbb{R}$ ($U(R(t)) : (R(t), S(t), A(t), H) \rightarrow R(t+1)$)
or
  $R : S \times A \rightarrow \mathbb{R}$
or
  $R : H \rightarrow \mathbb{R}$ (fitness) ($U(R(t)) : (R(t), S(t), A(t), H) \rightarrow R(t+1)$)
where one has typically $H = A(0), \ldots, A(N)$
with a behavioural function $B : \pi_\theta \rightarrow H$
Arising from a developmental function $D : \theta \rightarrow \pi_\theta$

• An optimization procedure (evolutionary algorithms or model learning with approximate dynamic programming/stochastic optimal control) that find a policy such that
  $\theta = \arg\max_{\theta_i} R(B(D(\theta_i)))$ (evo systems)

  $A(t) = \arg\max_{a \in A} \sum_{n=t+1}^{\infty} \gamma^t R(n)$ (devo-learning systems)
A common underlying formulation

• A (discrete or continuous) state space \( S \) (e.g. sensori state and memory of a robot)
• A (discrete or continuous) action space \( A \) (e.g. motor commands of a robot)
• A transition function \( W : S(t) \times A(t) \rightarrow S(t+1) \)
• A parameterized action policy \( \pi_\theta : S \rightarrow A \)

• A reward/value/fitness function
  \[ R : S \rightarrow \mathbb{R} \ (U(R(t)) : (R(t), S(t), A(t), H) \]
  or \( R : S \times A \rightarrow \mathbb{R} \)
  or \( R : H \rightarrow \mathbb{R} \) (fitness) \( U(R(t)) : (R(t), S(t), A(t), H) \)
  where one has typically \( H = A(0), ..., A(N) \)
  with a behavioural function \( B : \pi_\theta \rightarrow H \)
  Arising from a developmental function \( D : \theta \)

• An optimization procedure (evolutionary algorithms or model learning with approximate dynamic programming/stochastic optimal control) that find a policy such that
  \[ \theta = \arg \max_{\theta_i} R(B(D(\theta_i))) \]  (evo systems)

  \[ A(t) = \arg \max_{a \in A} \sum_{n=t+1}^{\infty} \gamma^n R(n) \]  (devo-learning systems)

Various research focuses:

i) Optimization mechanisms
ii) Policy representations
iii) S, A, D, B representations
iv) World model learning
v) World structure
vi) Reward/fitness function
Examples of engineering/task-specific evolutionary problem

\[ R(B(D(\theta))) = R(B(\pi_\theta)) = \text{fitness}(\theta) \]

= performance in tactile object recognition

\[ \text{fitness}(\theta) = \text{forward speed of robot} \]


Examples of engineering/task-specific developmental problem

\[ R(S(t), A(t)) = \text{is the ball in the cup?} \]

\[ R(S(t), A(t)) = \text{forward speed of the robot} \]


Example of modeling work for understanding the origins of open-ended complexity/structures in evolutionary systems

\[ \text{fitness}(\theta) = \text{number of offsprings/food collected} \]

e.g. Tierra (Ray, 1992), Polyworld (Yaeger, 1994), Geb (Channon, 2001)

Evolution of complexity of neural structures and behaviours (e.g. modularization, abstractions, cooperation...)

Evolution of structures at the population/ecology level (e.g. speciation, temporal dynamics...)


Example of modeling work for understanding the origins of open-ended complexity/structures in developmental systems

$R(S(t), A(t)) = \text{local derivative of certainty of learnt model of } S \times A \rightarrow S \text{ in the vicinity of } S(t)$

Self-organization of developmental patterns

Regularity AND diversity

<table>
<thead>
<tr>
<th>Measure</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure 1 (number of peaks?)</td>
<td>9.67</td>
</tr>
<tr>
<td>Measure 2 (complete scenario?)</td>
<td>Yes: 34 %, No: 66 %</td>
</tr>
<tr>
<td>Measure 3 (near complete scenario?)</td>
<td>Yes: 53 %, No: 47 %</td>
</tr>
<tr>
<td>Measure 4 (non-affordant bite before affordant bite?)</td>
<td>Yes: 93 %, No: 7 %</td>
</tr>
<tr>
<td>Measure 5 (non-affordant bash before affordant bash?)</td>
<td>Yes: 57 %, No: 43 %</td>
</tr>
<tr>
<td>Measure 6 (period of systematic successful bite?)</td>
<td>Yes: 100 %, No: 0 %</td>
</tr>
<tr>
<td>Measure 7 (period of systematic successful bash?)</td>
<td>Yes: 78 %, No: 11 %</td>
</tr>
<tr>
<td>Measure 8 (bite before bash?)</td>
<td>Yes: 92 %, No: 8 %</td>
</tr>
<tr>
<td>Measure 9 (successful bite before successful bash?)</td>
<td>Yes: 77 %, No: 23 %</td>
</tr>
</tbody>
</table>
Open-ended search for solving more efficiently engineering problems in developmental systems

Combining problem-specific rewards and intrinsic rewards for solving more efficiently specific problems

\[ R(S(t), A(t)) = \text{monkey crying} + \text{prediction errors of learnt model of } S \times A \rightarrow S \text{ in the vicinity of } S(t) \]

Efficient active model learning

Open-ended search for solving more efficiently engineering problems in evolutionary systems

Multi-objective evolutionary optimization (Pareto front based):

\[ \text{fitness}(\theta) = \text{distance to goal AND (phenotypic or behavioural) novelty} \]


Evolutionary modeling of the origins of intrinsic rewards

For every agent $A$, there is a space of reward functions $R_A$ that maps features of the history of observation-action pairs to scalar primary reward values (the specific choice of features is determined in defining $A$). There is a distribution over sequential decision making environments in some set $E$ in which we want our agent to perform well (in expectation). A specific reward function $r_A \in R_A$ and a sampled environment $E \in E$ produces $h$, the history of agent $A$ adapting to environment $E$ over its lifetime using the reward function $r_A$, i.e., $h \sim \langle A(r_A), E \rangle$, where $\langle A(r_A), E \rangle$ makes explicit that agent $A$ is using reward function $r_A$ to interact with environment $E$ and $h \sim \langle \cdot \rangle$ makes explicit that history $h$ is sampled from the distribution produced by the interaction $\langle \cdot \rangle$. A given fitness function $F$ produces a scalar evaluation $F(h)$ for all such histories $h$. An optimal reward function $r_A^* \in R_A$ is the reward function that maximizes the expected fitness over the distribution of environments, i.e.,

$$r_A^* = \arg \max_{r_A \in R_A} \mathbb{E}_{E \in E} \{F(h) | h \sim \langle A(r_A), E \rangle \}. \quad (1)$$

Evolutionary search over a reward functions in a reward space

An evo-devo-learning system!

$$\text{fitness}(\theta) = \text{fitness(learnt policy}|\text{reward}(\theta)) =$$

Result: optimal rewards do not directly correspond to distance to the out door, but includes explicit reward for novelty