Evolvability in Learning Theory

Dimitris Diochnos

Department of Computer Science
University of Virginia

April 15, 2019
Charlottesville, VA
Evolvability [Valiant, 2009] was based on Darwin’s work *On the Origin of Species by Means of Natural Selection* [Darwin, 1859].
Evolvability

Key Points

- Species *(Hypotheses)*, Generations *(Iterations)*.
- A *fitness* function called *performance*.
  - Estimated through *sampling*.
- Mutations define the Neighborhood.
- **Tolerance** \( t \) *partitions* the Neighborhood:
  - *Bene* = \( \{ h' \mid \text{Perf}_{D_n} (h', c) > \text{Perf}_{D_n} (h, c) + t \} \).
  - *Neut* = \( \{ h' \mid \text{Perf}_{D_n} (h', c) \geq \text{Perf}_{D_n} (h, c) - t \} \setminus \text{Bene} \).
  - *Deleterious*, the rest.

Goal

\[ \Pr (\text{Perf}_{D_n} (h, c) < \text{Perf}_{D_n} (c, c) - \varepsilon) < \delta. \] \hspace{1cm} (1)

Evolution should proceed from any starting point!
The Swapping Algorithm on Monotone Conjunctions

$t = 0.1$

$q$

$x \land y$

0.5
The Swapping Algorithm on Monotone Conjunctions

\[ x \land y = 0.5 \]

\[ t = 0.1 \]

\[ x = 0.25 \]
\[ y = 0 \]

\[ q \]
The Swapping Algorithm on Monotone Conjunctions

t = 0.1

\( x \land y \)
0.5

\( x \land y \land z \)
0.88

\( x \land y \land w \)
0.69
The Swapping Algorithm on Monotone Conjunctions

t = 0.1

\[ x \land y \]
\[ 0.5 \]

\[ x \]
\[ 0.25 \]

\[ y \]
\[ 0 \]

\[ x \land y \]
\[ 0.5 \]

\[ x \land z \]
\[ 0.63 \]

\[ w \land y \]
\[ 0.44 \]

\[ x \land y \land z \]
\[ 0.88 \]

\[ x \land y \land w \]
\[ 0.69 \]
Performance

\[ X_n = \{0, 1\}^n. \]
\[ h(x), c(x) \in \{+1, -1\}. \]

\[
\text{Perf}_{D_n} (h, c) = \sum_{x \in X_n} h(x) c(x) D_n(x) \\
= 1 - 2 \cdot \Pr (h(x) \neq c(x)) \\
= \mathbb{E} [h \cdot c].
\]

\[ \text{Estimated through sampling}, \]
\[ \text{Perf}_{D_n} (h, c, S) = \frac{1}{|S|} \sum_{x \in S} h(x) \cdot c(x). \]
Preliminary Remarks

Remark 1 (vs. PAC)

_Evolvability is a restricted case of PAC learnability._

Goal 1 (Evolvability)

\[
\Pr (\text{Perf}_{D_n} (h, c) < \text{Perf}_{D_n} (c, c) - \varepsilon) < \delta .
\]

Goal 2 (PAC Learning)

\[
\Pr (\text{error}_{D_n} (h, c) > \varepsilon) < \delta .
\]
Preliminary Remarks

Remark 2 (on the *Updates*)

*Updates depend only on the positivity and negativity of the examples or experiences, in the sense that there is no dependence on the description of the examples* (as is the case in the Statistical Query model); e.g., # of 1’s in binary representation.

Remark 3 (vs. *SQ* model, Valiant, 2009)

*Evolvable function classes $\subset$ SQ learnable function classes.*
Preliminary Remarks

Description 1 (The Tool on the SQ Model is a Query)

▷ Let $\psi : \{0, 1\}^n \times \{-1, 1\} \mapsto \{-1, 1\}$.
▷ A query is a pair $(\psi, \tau)$.
▷ Estimate $\mathbb{E} [\psi(x, \ell)]$ within tolerance $\tau$.

Description 2 (Types of Queries)

▷ independent of the target (i.e. $\psi$ depends only on $x$)
▷ correlational if $\psi(x, \ell) \equiv g(x)c(x)$.

Proposition 1
Any statistical query can be substituted by two statistical queries that are independent of the target and two correlational queries.
A Simulation Result

Remark 4 (CSQ Learnability $\Rightarrow$ Evolvability; Feldman 2008)

Let $\mathcal{C}$ be a concept class CSQ learnable over a class of distributions $\mathcal{D}$ by a polynomial time algorithm $A$. Then, there exists an evolutionary algorithm $N(A)$ such that $\mathcal{C}$ is evolvable by $N(A)$ over $\mathcal{D}$. 
Related Results in Evolvability

Feldman
- CSQ → Evolvability algorithm [Feldman, 2008].
- Full conjunctions are evolvable [Feldman, 2009].
- Monotone conjunctions are not evolvable distribution-independently using Boolean loss [Feldman, 2011].
- Monotone conjunctions are evolvable distribution-independently using quadratic loss [Feldman, 2012].

D, Turán and D
- Swapping algorithm under $\mathcal{U}_n$ [DT, 2009].
- Swapping algorithm under any $\mathcal{B}_n$ [D, 2016].
- (1+1) EA under some $\mathcal{B}_n$ [D, under submission].

Kanade, Valiant, Vaughan
- Evolvability with drifting targets [KVV, 2010]. (To be presented on April 29)

Kanade
- Recombination, parallel CSQ learning and general conjunctions [Kanade, 2011].

More Results
Basic Notation

Representation

- Hypotheses are conjunctions of boolean variables; e.g., \( h_1 = x_1 \land x_5 \land x_8 \).
- Size / length: \# vars in the conjunction; e.g., \( |h_1| = 3 \).
- Represented as a set of indices; e.g., \( h_1 = \{1, 5, 8\} \).
- Also useful: represented by a bitstring; e.g., \( h_1 = 10001001 \).
- Hamming distance \( d(h_1, h_2) \): \# positions where the bitstrings representing \( h_1 \) and \( h_2 \) differ.

Hypothesis Space

\( \mathcal{H} = C_n^{\le q} \). Hypotheses such that \( 0 \le |h| \le q \). (← non-realizable)

\( \mathcal{H} = C_n = C_n^{\le q} \cup C_n^{> q} \). Hypotheses such that \( 0 \le |h| \le n \).
Concept Class and Hypothesis Space

The diagram illustrates the concept class and hypothesis space for a given number of variables, $x_1, x_2, \ldots, x_n$. The level $n$ represents all possible conjunctions with precisely $q$ variables. The levels $q, 2, 1, 0$ correspond to conjunctions with $q$, $2$, $1$, and $0$ variables, respectively. The hypothesis space encompasses all possible conjunctions within these levels.
Monotone Conjunctions under the Uniform Distribution are Evolvable

<table>
<thead>
<tr>
<th>properties</th>
<th>[Valiant, 2007]</th>
<th>[D &amp; Turán, 2009]</th>
<th>[D, 2016]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{H} = C_n$</td>
<td>$\mathcal{H} = C_n$</td>
<td>$\mathcal{H} = C_n^{\leq q}$</td>
<td></td>
</tr>
<tr>
<td>$q$</td>
<td>$\mathcal{O}(\lg(n/\varepsilon))$</td>
<td>$\mathcal{O}(\lg(1/\varepsilon))$</td>
<td>$\mathcal{O}(\lg(1/\varepsilon))$</td>
</tr>
<tr>
<td>generations</td>
<td>$\mathcal{O}(n\lg(n/\varepsilon))$</td>
<td>$\mathcal{O}(n\lg(1/\varepsilon))$</td>
<td>$2q$</td>
</tr>
<tr>
<td>sample size</td>
<td>$\tilde{\mathcal{O}}((n/\varepsilon)^6)$</td>
<td>$\tilde{\mathcal{O}}(n^2/\varepsilon^2 + n/\varepsilon^4)$</td>
<td>$\tilde{\mathcal{O}}(n/\varepsilon^4)$</td>
</tr>
</tbody>
</table>

Theorem 1 (D & Turán, 2009)

Set $q = \lceil \lg(3/\varepsilon) \rceil$. For every target conjunction $c$ and every initial hypothesis $h_0$ it holds that after $\mathcal{O}(q + |h_0| \ln \frac{1}{\delta})$ iterations, each iteration evaluating the performance of $\mathcal{O}(nq)$ hypotheses, and each performance being evaluated using sample size $\mathcal{O}\left(\left(\frac{1}{\varepsilon}\right)^4 \left(\ln n + \ln \frac{1}{\delta} + \ln \frac{1}{\varepsilon}\right)\right)$ per iteration, the goal is achieved.
Correlation under the Uniform Distribution

\[ h = \bigwedge_{i \in M} x_i \land \bigwedge_{\ell \in \mathcal{M}} x_{\ell} \quad \text{and} \quad c = \bigwedge_{i \in M} x_i \land \bigwedge_{k \in \mathcal{U}} x_k \quad (2) \]

\[
\text{Perf}_{\mathcal{U}_n}(h, c) = 1 - 2^{1-(m+u)} - 2^{1-(m+r)} + 2^{2-(m+r+u)} \\
= 1 - 2^{1-|c|} - 2^{1-|h|} + 2^{2-|h|-u}
\]
Strategy

\[ h = \bigwedge_{i \in M} x_i \land \bigwedge_{\ell \in N} x_\ell \quad \text{and} \quad c = \bigwedge_{i \in M} x_i \land \bigwedge_{k \in U} x_k \]

- Short target \( \Rightarrow \) Find target precisely (w.h.p.)
- Long target \( \Rightarrow \) Find some good approximation (w.h.p.)
Strategy

\[ h = \bigwedge_{i \in M} x_i \wedge \bigwedge_{\ell \in R} x_\ell \quad \text{and} \quad c = \bigwedge_{i \in M} x_i \wedge \bigwedge_{k \in U} x_k \]

- Short target \( \Rightarrow \) Find target precisely (w.h.p.)
- Long target \( \Rightarrow \) Find some good approximation (w.h.p.)

**Lemma 2 (Performance Lower Bound)**

If \( |h| \geq q \) and \( |c| \geq q + 1 \) then \( \text{Perf}_{\mathcal{U}_n}(h, c) > 1 - 3 \cdot 2^{-q} \).

**Corollary 3**

Let \( q \geq \lg(3/\varepsilon), \ |h| \geq q, \ |c| \geq q + 1 \) \( \Rightarrow \) \( \text{Perf}_{\mathcal{U}_n}(h, c) > 1 - \varepsilon \).
Guiding the Search

\[ \Delta = \text{Perf}_{\mathcal{U}_n} (h', c) - \text{Perf}_{\mathcal{U}_n} (h, c) \]

**Theorem 4 (Structure of Best Approximations)**

The best q-approximation of a target c is

- c itself if \(|c| \leq q\)
- any hypothesis formed by q good variables if \(|c| > q\).
Example 1: Short Initial Hypothesis and Short Target

Let $X_8 = \{0, 1\}^8$ such that $\{g_1, g_2, g_3, b_1, b_2, b_3, b_4, b_5\}$, the target be $c = g_1 \land g_2 \land g_3$, and require $\varepsilon = 1/5$. 

\[
\begin{array}{cccc|c|c}
\text{Step } i & u & \text{Hypothesis } h_i & \text{Performance} & \text{Neighborhood} & \text{Class} \\
0 & 0 & \emptyset & -\frac{3}{4} & N^+ & \\
1 & 1 & \{b_1\} & 0 & N^+ \cup \{\text{swaps: } b \rightarrow g\} & \\
2 & \geq 2 & \{b_1 \land b_2\} & \frac{3}{8} & N^+ \cup \{\text{swaps: } b \rightarrow g\} & \\
3 & \geq 2 & \{b_1 \land b_2 \land b_3\} & \frac{9}{16} & N^+ \cup \{\text{swaps: } b \rightarrow g\} & \\
4 & \geq 2 & \{b_1 \land b_2 \land b_3 \land b_4\} & \frac{21}{32} & \{\text{swaps: } b \rightarrow g\} & \\
5 & \geq 2 & \{b_1 \land g_3 \land b_3 \land b_4\} & \frac{22}{32} & \{\text{swaps: } b \rightarrow g\} & \\
6 & 1 & \{g_1 \land g_3 \land b_3 \land b_4\} & \frac{24}{32} & \{\text{swaps: } b \rightarrow g\} & \\
7 & 0 & \{g_1 \land g_3 \land g_2 \land b_4\} & \frac{28}{32} & \{\text{remove } b\} & \\
8 & 0 & \{g_1 \land g_3 \land g_2\} & 1 & \{h_8\} & \\
\end{array}
\]
Example 2: Short Initial Hypothesis and Long Target

Let \( X_{13} = \{0, 1\}^{13} \) such that 
\( \{g_1, g_2, g_3, g_4, g_5, g_6, g_7, b_1, b_2, b_3, b_4, b_5, b_6\} \), the target be 
\( c = g_1 \land g_2 \land g_3 \land g_4 \land g_5 \land g_6 \land g_7 \), and require \( \varepsilon = 1/5 \). \( (q = 4) \)

<table>
<thead>
<tr>
<th>Step ( i )</th>
<th>( u )</th>
<th>Hypothesis ( h_i )</th>
<th>Performance</th>
<th>Neighborhood</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>( \geq 2 )</td>
<td>( \emptyset )</td>
<td>(-63/64)</td>
<td>( N^+ )</td>
<td>Bene</td>
</tr>
<tr>
<td>1</td>
<td>( \geq 2 )</td>
<td>( b_1 )</td>
<td>0</td>
<td>( N^+ \cup {\text{swaps: } b \rightarrow g} )</td>
<td>Bene</td>
</tr>
<tr>
<td>2</td>
<td>( \geq 2 )</td>
<td>( b_1 \land b_2 )</td>
<td>( 63/128 )</td>
<td>( N^+ \cup {\text{swaps: } b \rightarrow g} )</td>
<td>Neut</td>
</tr>
<tr>
<td>3</td>
<td>( \geq 2 )</td>
<td>( b_1 \land b_2 \land b_3 )</td>
<td>( 189/256 )</td>
<td>( N^+ \cup {\text{swaps: } b \rightarrow g} )</td>
<td>Neut</td>
</tr>
<tr>
<td>4</td>
<td>( \geq 2 )</td>
<td>( b_1 \land b_2 \land b_3 \land b_4 )</td>
<td>( 425/512 )</td>
<td>{all swaps} \cup {h_4}</td>
<td>Neut</td>
</tr>
<tr>
<td>5</td>
<td>( \geq 2 )</td>
<td>( b_1 \land b_6 \land b_3 \land b_4 )</td>
<td>( 425/512 )</td>
<td>{all swaps} \cup {h_5}</td>
<td>Neut</td>
</tr>
<tr>
<td>6</td>
<td>( \geq 2 )</td>
<td>( b_1 \land b_6 \land b_3 \land b_5 )</td>
<td>( 425/512 )</td>
<td>{all swaps} \cup {h_6}</td>
<td>Neut</td>
</tr>
<tr>
<td>7</td>
<td>( \geq 2 )</td>
<td>( b_1 \land b_6 \land b_3 \land b_5 )</td>
<td>( 425/512 )</td>
<td>{all swaps} \cup {h_7}</td>
<td>Neut</td>
</tr>
<tr>
<td>8</td>
<td>( \geq 2 )</td>
<td>( g_1 \land b_6 \land b_3 \land b_5 )</td>
<td>( 426/512 )</td>
<td>{swaps: ( b \rightarrow g)}</td>
<td>Bene</td>
</tr>
<tr>
<td>9</td>
<td>( \geq 2 )</td>
<td>( g_1 \land b_6 \land b_3 \land g_4 )</td>
<td>( 428/512 )</td>
<td>{swaps: ( b \rightarrow g)}</td>
<td>Bene</td>
</tr>
<tr>
<td>10</td>
<td>( \geq 2 )</td>
<td>( g_1 \land b_6 \land g_6 \land g_4 )</td>
<td>( 432/512 )</td>
<td>{swaps: ( b \rightarrow g)}</td>
<td>Bene</td>
</tr>
<tr>
<td>11</td>
<td>( \geq 2 )</td>
<td>( g_1 \land g_3 \land g_6 \land g_4 )</td>
<td>( 440/512 )</td>
<td>{swaps: ( g \rightarrow g)} \cup {h_{11}}</td>
<td>Neut</td>
</tr>
<tr>
<td>12</td>
<td>( \geq 2 )</td>
<td>( g_1 \land g_3 \land g_5 \land g_4 )</td>
<td>( 440/512 )</td>
<td>{swaps: ( g \rightarrow g)} \cup {h_{12}}</td>
<td>Neut</td>
</tr>
<tr>
<td>13</td>
<td>( \geq 2 )</td>
<td>( g_1 \land g_3 \land g_5 \land g_4 )</td>
<td>( 440/512 )</td>
<td>{swaps: ( g \rightarrow g)} \cup {h_{13}}</td>
<td>Neut</td>
</tr>
<tr>
<td>14</td>
<td>( \geq 2 )</td>
<td>( g_2 \land g_3 \land g_5 \land g_4 )</td>
<td>( 440/512 )</td>
<td>{swaps: ( g \rightarrow g)} \cup {h_{14}}</td>
<td>Neut</td>
</tr>
</tbody>
</table>